

Tilburg University

Optimal Pseudo-Gaussian and Rank-based Tests of the Cointegration Rank in Semiparametric Error-correction Models

Hallin, M.; Werker, B.J.M.; van den Akker, R.

Publication date:
2015

Document Version
Early version, also known as pre-print

[Link to publication in Tilburg University Research Portal](#)

Citation for published version (APA):

Hallin, M., Werker, B. J. M., & van den Akker, R. (2015). *Optimal Pseudo-Gaussian and Rank-based Tests of the Cointegration Rank in Semiparametric Error-correction Models*. (CentER Discussion Paper; Vol. 2015-001). CentER, Center for Economic Research.

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

No. 2015-001

**OPTIMAL PSEUDO-GAUSSIAN AND RANK-BASED
TESTS OF THE COINTEGRATION RANK IN
SEMIPARAMETRIC ERROR-CORRECTION MODELS**

By

Marc Hallin, Ramon van den Akker,
Bas J.M. Werker

8 January, 2015

ISSN 0924-7815
ISSN 2213-9532

Optimal pseudo-Gaussian and rank-based tests of the cointegration rank in semiparametric error-correction models

Marc Hallin^{a,b,c,1,*}, Ramon van den Akker^c, Bas J.M. Werker^{c,d}

^a*ECARES, Université Libre de Bruxelles*

^b*ORFE, Princeton University*

^c*Econometrics group, CentER, Tilburg University*

^d*Finance group, CentER, Tilburg University*

Abstract

This paper provides locally optimal pseudo-Gaussian and rank-based tests for the cointegration rank in linear cointegrated error-correction models with i.i.d. elliptical innovations. The proposed tests are asymptotically distribution-free, hence their validity does not depend on the actual distribution of the innovations. The proposed rank-based tests depend on the choice of scores, associated with a reference density that can freely be chosen. Under appropriate choices they are achieving the semiparametric efficiency bounds; when based on Gaussian scores, they moreover uniformly dominate their pseudo-Gaussian counterparts. Simulations show that the asymptotic analysis provides an accurate approximation to finite-sample behavior. The theoretical results are based on a complete picture of the asymptotic statistical structure of the model under consideration.

Keywords: Cointegration model, cointegration rank, elliptical densities, error-correction model, Lagrange multiplier test, Local Asymptotic Brownian Functional, Local Asymptotic Mixed Normality, Local Asymptotic Normality, multivariate ranks, quasi-likelihood procedures.

JEL codes: C14, C32.

*Corresponding author

Email addresses: mhallin@ulb.ac.be (Marc Hallin),
R.vdnAkker@TilburgUniversity.edu (Ramon van den Akker),
Werker@TilburgUniversity.edu (Bas J.M. Werker)

¹Royal Academy of Belgium. Supported by the IAP-network P7/06, Belgian Science Policy Office.

The authors wish to thank Peter Boswijk for pointing out a clever local parameterization of the cointegration model.

1. Introduction

Since their introduction in Granger (1981) and Engle and Granger (1987), cointegration models and the corresponding inference techniques have developed into a central topic in time-series econometrics, generating an extensive literature. The inferential side of that literature mainly deals with Gaussian, pseudo/quasi-Gaussian likelihood, or moment-based methods for problems related, e.g., to the cointegration rank or the cointegrating vectors; see, among many others, Stock (1987), Johansen and Juselius (1990), Johansen (1988, 1991, 1995), Phillips (1991), and Reinsel and Ahn (1992).

Whenever optimality issues—of a local and asymptotic nature in this context—are to be addressed, the adequate tool is Le Cam’s asymptotic theory of statistical experiments; see, e.g., Strasser (1985), Le Cam (1986), Le Cam and Yang (1990), or Van der Vaart (2000). The concept of *limit experiment*—more precisely, *limits of local sequences of experiments*—there plays an essential role: depending on their nature, those limit experiments indeed determine the asymptotic performances of tests and estimators, and the various efficiency bounds (parametric or nonparametric) that can be achieved. Often, they also suggest how to construct optimal procedures. That approach, for cointegration models, has been taken by several authors, including Phillips (1991), Jegannathan (1997) and Hodgson (1998a, 1998b), and exploited to construct optimal tests for hypotheses on the cointegration vectors.

This paper focuses on the construction of optimal tests for hypotheses on the cointegration rank in error-correction models (ECMs). We allow for possible deterministic linear time trends—generated by non-zero values of the parameter μ in model (2.1) below. The presence of such trends indeed has a dramatic impact on the nature of the various limit experiments. It leads, for specific directions, both within and outside the cointegrating space, to the familiar Locally Asymptotically Normal (LAN) structure, albeit with nonstandard convergence rate $T^{3/2}$ (Corollary 3.1). All possible limit experiments are characterized, in Proposition A.2, for non-seasonal cointegrated ECMs with independent and

identically elliptically distributed innovations. These limit experiments are generally of the complicated Locally Asymptotically Brownian Functional (LABF) type (Jeganathan (1995)). Considering, as a first step, the LAN subexperiment associated with perturbations of the cointegration rank only (no nuisances: all other parameters—cointegrating vectors and short-term dynamics—are supposed to be known), we construct new tests that are locally and asymptotically optimal (most stringent) for the cointegration rank, under specified innovation density (Section 3.2). Invoking adaptivity arguments, we then show (Section 4) that those tests actually remain optimal when all other parameters are treated as nuisances to be estimated—that is, in the full experiment (still, under specified innovation density). The tests turn out to be of the Lagrange Multiplier type.

The incentive for including possible trends in the model actually originates in applications, and many empirical studies long ago have incorporated this possibility in their analyses. This is the case, for instance, of Bernard and Durlauf (1995) in their study of convergence and common trends in per capita output (see their Equation (3)). In the area of asset pricing, Nasseh and Strauss (2000) explicitly allow for the presence of deterministic time trends when studying the relation between stock prices and macroeconomic activity (see their Equation (3)). Swift (2011) documents a long-run relationship between health and GDP in OECD countries; model (1) in that paper explicitly allows for a parameter μ generating linear time trends. A more recent example is Wong, Chiu and Wong (2014), in a study of optimal investment with longevity risks (see their Equations (4)–(5)). When present, trends can and should be exploited, with huge potential benefits. It should be insisted, though, that, while the optimality properties of our tests very much depend on their presence, validity (in terms of asymptotic size) remains unaffected by their absence.

Now, the actual underlying density in most applications remains unspecified, while the optimal parametric tests described in Section 3.2 typically lose their (asymptotic) validity under misspecified innovation densities. Pseudo-Gaussian (sometimes called Quasi Gaussian Maximum Likelihood, QMLE) methods then

are the common practice, therefore we start (Proposition 4.1) with deriving pseudo-Gaussian versions of the optimal parametric tests of Section 3.2. Those pseudo-Gaussian tests are quite satisfactory² when actual densities are close to Gaussian ones. This, however, (due, e.g., to heavy tails) needs not be the case; and pseudo-Gaussian methods unfortunately may exhibit rather poor performances away from the Gaussian. Traditional semiparametric methods (in the Bickel et al. (1993) style) in principle provide the semiparametrically optimal solution in such cases. But they remain theoretically and numerically quite heavy, as they require guessing appropriate tangent space projections (unless the problem is *adaptive*), running kernel estimation of innovation densities, usually with sample splitting, etc. We propose avoiding this by turning to rank-based techniques.

General results by Hallin and Werker (2003) indeed indicate that rank-based techniques offer an effective and numerically more tractable alternative, achieving semiparametric efficiency at chosen, (possibly, via data-driven methods) reference densities. Accordingly, we introduce, in Section 4.3, a class of test statistics involving a multivariate notion of residual signed ranks. Those statistics are obtained by projecting the optimal parametric test statistics of Section 3.2 onto those ranks, which in practice is extremely easy. The use of ranks is facilitated by the underlying assumption of elliptically distributed innovations, an assumption that has been considered in a number of contributions, see Hodgson et al. (2002) and Hodgson and Vorkink (2003) for two examples.

We then show that the rank-based versions of the locally and asymptotically most stringent tests associated with the reference density still achieve parametric optimality under that reference density—while remaining valid under the actual innovation density. Such rank-based tests offer several advantages. First of all, they are asymptotically distribution-free, so that their asymptotic critical values do not depend on the actual distribution of the innovations; were it not for the

²They are not admissible, though, being uniformly dominated by the van der Waerden version of our rank-based tests: see Section 4.4.

presence of estimated nuisance parameters, this distribution-freeness property would even hold exactly in finite samples. Second, while reaching parametric optimality under the reference density (which is not necessarily Gaussian), they often outperform, away from the reference density, sometimes quite significantly, the pseudo-Gaussian tests. In particular, the rank-based procedures associated with the Gaussian reference density (van der Waerden, or Gaussian-score tests) uniformly improve over the pseudo-Gaussian ones (see Section 4.4). In general, ranks³ provide a form of robustness that stabilizes finite-sample sizes (see Section 5). The use of ranks relies on the assumed elliptical error distribution. Pseudo-Gaussian procedures do not require that assumption for validity. The present paper thus quantifies the gains possible in applications where ellipticity is likely to hold.

The remainder of this paper is organized as follows. Section 2 gives a precise description of the model and model assumptions considered throughout. Section 3 presents a special case (relevant for testing hypotheses on the cointegration rank) of our general result on the (local) limit experiments resulting from the cointegrated error-correction model (the general result is available in Appendix A). Those limit experiments, for given innovation density, settle the parametric efficiency bounds, and provide the parametrically optimal procedures, for testing hypotheses on the cointegration rank. Exploiting the specific form of those parametrically optimal procedures, we provide, in Section 4, quasi-Gaussian and rank-based versions of the same tests, and study their properties under the null and local alternatives. In particular, we pay attention to the impact of computing ranks from estimated residuals instead of the actual innovations. These results are subsequently used in Proposition 4.3 to construct asymptotically distribution-free most-stringent rank-based tests of the cointegration rank; for the purpose of comparisons, we also provide (Propo-

³Note that ranks in the context of cointegration have been used before by Breitung and Gouriéroux (1997) and Breitung (2001)—in a totally different spirit (no optimality concerns), though, and based on a totally different concept of ranks.

sition 4.1) the explicit form of the corresponding optimal quasi-Gaussian tests. Even though the rates of convergence of our procedures are $T^{3/2}$, their finite sample performance is ultimately of interest. Section 5, therefore, provides a simulation study that shows that our asymptotic analysis indeed provides a most decent approximation to actual finite-sample performances and the differences in limiting power of the various tests are visible in finite samples. Further Monte Carlo results are provided in Appendix D. Technical results and proofs are concentrated in Appendices A–C.

2. The model

We consider realizations $X^{(T)} := (X_1, \dots, X_T)'$ from a p -dimensional stochastic process $\{X_t \mid t \in \mathbb{N}\}$ generated by the k -th order vector autoregressive model written in *error-correction form* (ECM)

$$\Delta X_t = \Pi X_{t-1} + \sum_{j=1}^{k-1} \Gamma_j \Delta X_{t-j} + \mu + \varepsilon_t, \quad t \in \mathbb{N}, \quad (2.1)$$

where Δ denotes first-order differencing, X_{1-k}, \dots, X_0 are deterministic starting values, $\Pi \in \mathbb{R}^{p \times p}$, $\Gamma := (\Gamma_1, \dots, \Gamma_{k-1}) \in \mathbb{R}^{p \times (k-1)p}$ and $\mu \in \mathbb{R}^p$ are parameters, and $\{\varepsilon_t\}$ is an i.i.d. sequence of elliptically distributed innovations (centered at the origin) with density \mathfrak{f} . We shall assume that $\{X_t\}$ is integrated of order one.

The assumptions we impose on this model are of two types: assumptions on the density \mathfrak{f} of ε_t (the innovation density), and assumptions on the parameters μ , Γ , and Π .

2.1. Innovation densities

We assume throughout that the innovations are elliptically distributed.

Assumption 1. (Elliptical symmetry) There exists a $p \times p$ symmetric positive definite matrix Σ and a function $f : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ satisfying $\int_0^\infty z^{p-1} f(z) dz = 1$

(the radial density) such that⁴

$$\mathfrak{f}(e) = \mathfrak{f}_{\Sigma, f}(e) := \frac{1}{\omega_p \sqrt{\det \Sigma}} f(\|e\|_{\Sigma}), \quad e \in \mathbb{R}^p, \quad (2.2)$$

where $\|e\|_{\Sigma} := (e' \Sigma^{-1} e)^{1/2}$ and ω_p denotes the Lebesgue measure of the unit sphere \mathcal{S}^{p-1} in \mathbb{R}^p . \square

Under Assumption 1, the *radial distance* $\|\varepsilon_t\|_{\Sigma}$ has density $\tilde{f}_p(z) := z^{p-1} f(z)$ at $z \in \mathbb{R}_+$; write \tilde{F}_p for the corresponding distribution function.⁵ On the radial density f we impose the following regularity assumption.

Assumption 2. (Radial density)

- (a) The radial density f is absolutely continuous with a.e. derivative f' , i.e. there exists a function f' such that

$$f(b) - f(a) = \int_a^b f'(z) dz \quad \text{for all} \quad 0 \leq a < b.$$

- (b) The radial Fisher information

$$\mathcal{I}_p(f) := \mathbb{E} \phi_f^2(\|\varepsilon_1\|_{\Sigma}) = \int_0^1 \phi_f^2(\tilde{F}_p^{-1}(u)) du,$$

where $\phi_f := -f'/f$ denotes the so-called location score of f , is finite.

- (c) $f(z) > 0$ for all $z \in \mathbb{R}_+$. \square

Observe that the location score for the p -variate density \mathfrak{f} is

$$-\text{grad}_e \log \mathfrak{f}(e) = \phi_f(\|e\|_{\Sigma}) \Sigma^{-1/2} u \quad \text{a.e., with } u := \Sigma^{-1/2} e / \|e\|_{\Sigma}, \quad e \in \mathbb{R}^p,$$

which explains our terminology for ϕ_f and $\mathcal{I}_p(f)$. It is possible to weaken Assumption 2 into an assumption of quadratic mean differentiability for the

⁴Clearly, (2.2) identifies f and Σ only up to a scale factor; requiring $\int_0^\infty z^{p-1} f(z) dz = 1$ thus only imposes an identification constraint, and does not restrict the model. Moreover, both Σ and f are nuisance parameters in our problem.

⁵Note that, contrary to \tilde{f}_p , and despite of its common name, the radial density f is not a probability density, as it does not integrate to one.

mapping $e \mapsto f^{1/2}(\|e\|)$; this is mainly of theoretical interest, though, and we refer to Section 1 of Hallin and Paindaveine (2002a) for details.

Let \mathcal{F} denote the set of all radial densities satisfying Assumption 2. The existence of moments for \mathbf{f} is completely determined by the existence of the corresponding moments for \tilde{f}_p ; denote by \mathcal{F}_2 the subset of \mathcal{F} yielding elliptical densities with finite second moment, i.e.,

$$\mathcal{F}_2 := \left\{ f \in \mathcal{F} \mid \int \zeta^2 \tilde{f}_p(\zeta) d\zeta = \int z^{p+1} f(z) dz < \infty \right\}.$$

2.2. Parameter restrictions

We are interested in the case that $\{X_t\}$ is integrated of order one, $I(1)$, and has no seasonal unit roots. The number of linearly independent cointegrating relationships, i.e. the dimension of the cointegration space, is denoted by r . The required restrictions on the parameters are well known (see, e.g., Johansen (1995)); to set notation, and for the sake of completeness, we briefly recall them here.

The *characteristic polynomial* $A_{\Gamma, \Pi}$ associated with (2.1) is

$$A_{\Gamma, \Pi}(z) := (1 - z)I_p - \Pi z - \sum_{j=1}^{k-1} \Gamma_j (1 - z)z^j, \quad z \in \mathbb{C}.$$

A (non-seasonal) unit root implies that Π is singular: indeed, we then have $0 = |A_{\Gamma, \Pi}(1)| = |\Pi|$, and Π has rank r , $0 \leq r \leq p - 1$. Accordingly, we can write Π as the product $\Pi = \alpha\beta'$ of two $p \times r$ matrices α and β of rank r ; in case $r = 0$, define $\alpha := 0_{p \times p} =: \beta$. Also, as usual, let α_\perp and β_\perp denote $p \times (p - r)$ matrices of rank $(p - r)$ satisfying $\alpha'\alpha_\perp = 0_{r \times (p-r)} = \beta'\beta_\perp$; for $r = 0$, define $\alpha_\perp := I_p =: \beta_\perp$.⁶

Using the notation above, the parameter restrictions are formalized as follows.

Assumption 3. The matrices $\Pi = \alpha\beta'$ and Γ in (2.1) are such that

- (a) the rank of Π is $r < p$;

⁶Note that the matrices $\alpha, \alpha_\perp, \beta$, and β_\perp are not uniquely defined (unless $r = 0$). That lack of identifiability plays no role in the sequel.

- (b) if $|A_{\Gamma,\Pi}(z)| = 0$ then $|z| > 1$ or $z = 1$;
- (c) the matrix $\Psi = \Psi_{\Gamma,\Pi} := \alpha'_\perp \left(I_p - \sum_{j=1}^{k-1} \Gamma_j \right) \beta_\perp$ is non-singular. \square

Assumption 3(b) excludes the possibility of explosive behavior and seasonal unit roots in the process $\{X_t\}$, and Assumption 3(c) is equivalent to the requirement that $z \mapsto A_{\Gamma,\Pi}(z)$ has exactly $(p - r)$ unit roots and prevents the process from being $I(2)$.⁷

2.3. Deterministic linear trend

Under Assumption 3, the following version of the Granger-Johansen representation theorem, see Hansen (2005) and Nielsen (2009), holds: the process $\{X_t\}$ generated by (2.1) admits the representation

$$X_t = \beta_\perp \Psi^{-1} \alpha'_\perp \sum_{s=1}^t (\varepsilon_s + \mu) + Y_t + a, \quad t \in \mathbb{N}, \quad (2.3)$$

where

- (i) a is a deterministic starting value satisfying $\beta' a = 0$;
- (ii) the process $\{Y_t\}$ is of the form $Y_t = \Upsilon V_t$ for some $p \times (r + (k - 1)p)$ matrix Υ , where

$$V_t' := ((\beta' X_t)', (\Delta X_t)', \dots, (\Delta X_{t-k+2})')$$

satisfies $V_t = c + \Xi V_{t-1} + \Omega \varepsilon_t$, $t \in \mathbb{N}$, for some $(r + (k - 1)p) \times p$ matrix Ω , $(r + (k - 1)p)$ -dimensional vector c , and $(r + (k - 1)p) \times (r + (k - 1)p)$ matrix Ξ whose eigenvalues are all less than one in absolute value.⁸

The representation (2.3) immediately implies that the cointegration vectors, i.e. the columns of β , eliminate both the deterministic linear trend and the stochastic common trends. As discussed in the introduction, we will focus on

⁷See, for example, Johansen (1995).

⁸The starting value a , the vector c and the matrices Υ , Ξ , and Ω all depend on the parameters $\mu, \Gamma_1, \dots, \Gamma_{k-1}$, α , and β . Exact formulas for these quantities are available (see Hansen (2005) and, in particular, Theorem 3.2 in Nielsen (2009)), but we do not need them for our purposes.

the situation in which at least one component of $\{X_t\}$ has a deterministic, linear time trend, i.e., at least one component of the p -vector $(\beta_\perp \Psi^{-1} \alpha'_\perp \mu)t$ does not vanish. Of course, this is only the case, under Assumption 3, if $\alpha'_\perp \mu \neq 0_{p-r}$. This motivates the following definition.

Definition 1. We say that the process $\{X_t\}$ defined in (2.1) contains a (deterministic and linear) time trend if $\alpha'_\perp \mu \neq 0_{p-r}$. \square

In Section 3, we derive the limit experiment of the cointegrated ECM (2.1) with a deterministic and linear time trend. We exploit this limit experiment to construct new statistics for testing hypotheses on the cointegration rank. While these tests are optimal (in a sense to be made precise later on) whenever $\alpha'_\perp \mu \neq 0_{p-r}$, we already stress that they remain valid (in the sense of correct asymptotic size) also in case none of the components has a linear deterministic time trend, i.e. $\alpha'_\perp \mu = 0_{p-r}$.

We conclude this section with some more notation. Let Θ denote the set of admissible values of the parameter $\vartheta := (\mu, \Gamma, \Pi)$, i.e., those values of ϑ where $\mu \in \mathbb{R}^p$ is unrestricted, but Γ and Π are such that Assumption 3 holds. We write $P_{\vartheta; \Sigma, f}^{(T)}$ or $P_{\mu, \Gamma, \Pi; \Sigma, f}^{(T)}$ for the distribution of the vector (X_1, \dots, X_T) generated by (2.1), conditional on the starting values X_{-k+1}, \dots, X_0 , under the Euclidean parameter values $(\vartheta; \Sigma) = (\mu, \Gamma, \Pi; \Sigma)$ and the infinite-dimensional parameter f .

3. Limit experiments and optimal inference in parametric submodels

A complete picture of the limiting local experiments associated with the ECM (2.1) is provided in the appendix in Proposition A.2. In this section, however, we focus on the particular case of that general result which is needed when performing inference about the cointegration rank. Section 3.1 deals with those particular limit experiments. Then, in Section 3.2, we exploit that result to study optimal inference on the cointegration rank in specific *parametric* submodels with specified innovation density. The form of the resulting optimal parametric tests then will help us deriving the *semiparametric* optimal ones we propose in Section 4.

3.1. Local limit experiments

To obtain the local limiting experiments, we analyze the (limiting) behavior of likelihood ratios for specific local perturbations $\vartheta^{(T)} := (\mu^{(T)}, \Gamma^{(T)}, \Pi^{(T)})$ of $\vartheta = (\mu, \Gamma, \Pi) \in \Theta$. Note that, since Θ is not an open set, it may happen that $\vartheta^{(T)}$ does not belong to Θ and possibly corresponds, for instance, to explosive-root alternatives. The results we are deriving nevertheless also hold for such alternatives. The model we are investigating is, in that sense, slightly larger than the one parametrized by Θ ; this is common practice in the cointegration literature—see, e.g. Johansen (1995) on Gaussian maximum likelihood estimators. We do not need to consider local alternatives for (Σ, f) as it turns out that the testing problems we consider in this paper are *adaptive* with respect to (Σ, f) , see Section 4.

Building on the factorization $\Pi = \alpha\beta'$, where α and β are full-rank $p \times r$ matrices, define local (here, local to $\vartheta = (\mu, \Gamma, \Pi)$) sequences of alternatives of the form $\vartheta^{(T)} = (\mu^{(T)}, \Gamma^{(T)}, \Pi^{(T)})$, with

$$\mu^{(T)} = \mu_m^{(T)} := \mu + T^{-1/2}m, \quad \Gamma^{(T)} = \Gamma_G^{(T)} := \Gamma + T^{-1/2}G, \quad (3.1)$$

and

$$\Pi^{(T)} = \Pi_{A,b,B,d}^{(T)} := \alpha_A^{(T)} \beta_{b,B}^{(T)'} + T^{-3/2} \alpha_{\perp} d (\beta_{\perp} \Psi^{-1} \alpha'_{\perp} \mu)', \quad (3.2)$$

where

$$\alpha^{(T)} = \alpha_A^{(T)} := \alpha + T^{-1/2}A \text{ and } \beta^{(T)} = \beta_{b,B}^{(T)} := \beta + T^{-3/2}(\beta_{\perp} \Psi^{-1} \alpha'_{\perp} \mu)b' + T^{-1}\beta_{\perp} B'. \quad (3.3)$$

The local parameters characterizing those alternatives are thus

$$m \in \mathbb{R}^p, G_1, \dots, G_{k-1} \in \mathbb{R}^{p \times p}, A \in \mathbb{R}^{p \times r}, b \in \mathbb{R}^r, B \in \mathbb{R}^{r \times p-r}, \text{ and } d \in \mathbb{R}^{p-r}.$$

Recall that α and β are not separately identified, as only their product $\Pi = \alpha\beta'$ is, which implies that the interpretation of the local parameters A , b , and B depends on the factorization adopted for Π and the chosen versions of α_{\perp} and β_{\perp} . This, however, has no further consequences in the sequel.

The localizing rates (the standard $T^{-1/2}$ and “super-consistency” rates T^{-1} and $T^{-3/2}$) are such that $\Pi_{A,b,B,d}^{(T)}$ yields a contiguous alternative to Π . Contiguous means that the induced probability measures are neither asymptotically orthogonal (such that testing becomes trivial), nor equal (such that testing becomes impossible). See, for example, Chapter 6 in Van der Vaart (2000) for a formal discussion. Note that perturbations at rate $T^{-3/2}$ are only detectable if the process contains a deterministic linear trend, i.e., when $\alpha'_\perp \mu \neq 0_{p-r}$. The perturbations of μ , Γ and α are standard, while the perturbation of Π is somewhat more subtle. The local parameters A , b , and B perturb the space of adjustment coefficients and modify the cointegrating space without affecting (for T large enough) the cointegration rank r . Provided that the process contains a deterministic linear trend (i.e., $\alpha'_\perp \mu \neq 0_{p-r}$), a perturbation $d \neq 0$ does (for T large enough) affect the cointegration rank by increasing it from r to $r + 1$.

To describe the limit experiments associated with these perturbations, we introduce the following notation. For all ϑ and $\Sigma > 0$, define the residuals

$$\epsilon_t = \epsilon_t(\vartheta) := \Delta X_t - \Pi X_{t-1} - \sum_{j=1}^{k-1} \Gamma_j \Delta X_{t-j} - \mu \quad (3.4)$$

and, with $\Sigma^{1/2}$ the symmetric square root of Σ , the corresponding *sphericized* unit vectors (playing the role of *multivariate signs*)

$$U_t = U_t(\vartheta; \Sigma) := \frac{1}{\|\epsilon_t(\vartheta)\|_\Sigma} \Sigma^{-1/2} \epsilon_t(\vartheta), \quad (3.5)$$

with the convention that $U_t := 0$ in case $\epsilon_t = 0$.

Under Assumption 1, $U_t^\varepsilon := \Sigma^{-1/2} \varepsilon_t / \|\varepsilon_t\|_\Sigma$ is uniformly distributed over the unit sphere \mathcal{S}^{p-1} and independent of $\|\varepsilon_t\|_\Sigma$. And, if ϑ and Σ are the true parameter values, that is, under $P_{\vartheta; \Sigma, f}^{(T)}$, $\epsilon_t(\vartheta)$ and $U_t(\vartheta; \Sigma)$ coincide with ε_t and U_t^ε , respectively. This explains the above terminology.

We are now ready to describe the limit experiments.

Proposition 3.1. *Consider the Error Correction Model (2.1) with a deterministic linear trend, i.e., assume $\alpha'_\perp \mu \neq 0_{p-r}$. Let Assumptions 1-3 hold, $\vartheta \in \Theta$,*

and $f \in \mathcal{F}_2$. Consider a bounded sequence of perturbations

$$h'_T := (m'_T, (\text{vec } G_T)', (\text{vec } A_T)', b'_T, (\text{vec } B_T)', d'_T),$$

with $\beta_\perp B'_T$ orthogonal to $\beta_\perp \Psi^{-1} \alpha'_\perp \mu$, defining a local parameter sequence $\vartheta^{(T)}$, see (3.1)-(3.3). Then, under $P_{\vartheta; \Sigma, f}^{(T)}$, as $T \rightarrow \infty$, we have

$$\log \frac{dP_{\vartheta^{(T)}; \Sigma, f}^{(T)}}{dP_{\vartheta; \Sigma, f}^{(T)}} = h'_T \Delta_\vartheta^{(T)} - \frac{1}{2} h'_T J_\vartheta^{(T)} h_T + o_P(1), \quad (3.6)$$

with central sequence⁹ $\Delta_\vartheta^{(T)} = (\Delta_\mu^{(T)'}, \Delta_\Gamma^{(T)'}, \Delta_\alpha^{(T)'}, \Delta_b^{(T)'}, \Delta_B^{(T)'}, \Delta_d^{(T)'})'$, where $\Delta_\Gamma^{(T)}$ is shorthand notation for $(\Delta_{\Gamma_1}^{(T)'}, \dots, \Delta_{\Gamma_{k-1}}^{(T)'})'$, given by

$$\Delta_\vartheta^{(T)} := \sum_{t=1}^T \begin{pmatrix} Z_{Tt}^{(1)}(\vartheta) \otimes I_p \\ Z_{Tt}^{(2)}(\vartheta) \otimes \alpha' \\ Z_{Tt}^{(2,1)}(\vartheta) \otimes \alpha'_\perp \end{pmatrix} \Sigma^{-1/2} U_t(\vartheta) \phi_f(\|\epsilon_t(\vartheta)\|_\Sigma), \quad (3.7)$$

and Fisher Information

$$J_\vartheta^{(T)} := \frac{\mathcal{I}_p(f)}{p} \sum_{t=1}^T \begin{pmatrix} Z_{Tt}^{(1)} Z_{Tt}^{(1)'} \otimes \Sigma^{-1} & Z_{Tt}^{(1)} Z_{Tt}^{(2)'} \otimes \Sigma^{-1} \alpha & Z_{Tt}^{(1)} Z_{Tt}^{(2,1)'} \otimes \Sigma^{-1} \alpha_\perp \\ & Z_{Tt}^{(2)} Z_{Tt}^{(2)'} \otimes \alpha' \Sigma^{-1} \alpha & Z_{Tt}^{(2)} Z_{Tt}^{(2,1)'} \otimes \alpha' \Sigma^{-1} \alpha_\perp \\ & & Z_{Tt}^{(2,1)} Z_{Tt}^{(2,1)'} \otimes \alpha'_\perp \Sigma^{-1} \alpha_\perp \end{pmatrix},$$

where $Z_{Tt}^{(1)} = Z_{Tt}^{(1)}(\vartheta)$ and $Z_{Tt}^{(2)} = Z_{Tt}^{(2)}(\vartheta)$ are defined by

$$Z_{Tt}^{(1)'} := T^{-1/2} \begin{pmatrix} 1 & \Delta X_{t-1} & \cdots & \Delta X_{t-k+1} & \beta' X_{t-1} \end{pmatrix}, \quad \text{and} \quad (3.8)$$

$$Z_{Tt}^{(2)} := \begin{pmatrix} Z_{Tt}^{(2,1)} \\ Z_{Tt}^{(2,2)} \end{pmatrix} = \begin{pmatrix} T^{-3/2} (\beta_\perp \Psi^{-1} \alpha'_\perp \mu)' X_{t-1} \\ T^{-1} \beta'_\perp (X_{t-1} - (t-1) \beta_\perp \Psi^{-1} \alpha'_\perp \mu) \end{pmatrix}, \quad (3.9)$$

respectively.

Moreover, still under $P_{\vartheta; \Sigma, f}^{(T)}$, $(\Delta_\vartheta^{(T)}, J_\vartheta^{(T)})$ converges in distribution¹⁰ to (Δ, J) with $E \exp(h' \Delta - h^2 J/2) = 1$ for all h . The distribution of (Δ, J) is provided in Appendix A. Denoting by Δ^- the subvector obtained from deleting, in Δ , the block of components corresponding to B_T , and by J^- the submatrix resulting from similarly deleting, in J , the corresponding rows and columns, J^- is deterministic, and $\Delta^- \sim N(0, J^-)$. \square

⁹ $A \otimes BC$ stands for $A \otimes (BC)$; see, e.g., Magnus and Neudecker (1988).

¹⁰ Throughout, we use the notation $(A_T, B_T) \xrightarrow{d} (A, B)$ as shorthand for $(\text{vec}(A_T)', \text{vec}(B_T)')' \xrightarrow{d} (\text{vec}(A)', \text{vec}(B)')$.

Proof. The result directly follows from Proposition A.2 in the appendix by letting $D_T = 0$. \square

This result calls for some remarks.

Remark 3.1.

- (i) As $\beta_\perp \Psi^{-1} \alpha'_\perp \mu$ lies in the column space of β_\perp , b_T and B_T are not separately identified. Therefore, imposing on B_T a linear constraint of the form $\beta_\perp B'_T \perp (\beta_\perp \Psi^{-1} \alpha'_\perp \mu)$ does not entail any loss of generality. This constraint actually illustrates the rationale for splitting the local perturbations of β into b_T and B_T . From (2.3) we see that, under $P_{\vartheta; \Sigma, f}$, $\beta' X_t$ is stationary and has zero drift (as $\beta' \beta_\perp \Psi^{-1} \alpha'_\perp \mu = 0$). For local perturbations $\beta^{(T)}$ of β induced (at rate T) by B_T , this still holds (under $P_{\vartheta; \Sigma, f}$): indeed, $\beta^{(T)'} X_t$ remains stationary since the additional effect of B_T on the drift is $T^{-1} B_T \beta'_\perp \beta_\perp \Psi^{-1} \alpha'_\perp \mu = 0$. On the other hand, the local perturbations $\beta^{(T)}$ of β induced (at rate $T^{3/2}$) by b_T lead, under $P_{\vartheta; \Sigma, f}$, and provided that $\beta_\perp \Psi^{-1} \alpha'_\perp \mu \neq 0$, to non-zero drifts in $\beta^{(T)'} X_t$, of magnitude $T^{-3/2} b_T |\beta_\perp \Psi^{-1} \alpha'_\perp \mu|^2$. These effects on drifts also explain the different localizing rates for b_T and B_T .
- (ii) While a formal derivation of the central sequence is given in the proof of Proposition A.2, the form of $\Delta_\vartheta^{(T)}$ in (3.7) also follows by pointwise differentiation of the log-likelihood. Such differentiation does not yield the terms $-(t-1) \beta_\perp \Psi^{-1} \alpha'_\perp \mu$ in $Z_{Tt}^{(2,2)}(\vartheta)$. However, these terms, corresponding to B_T , vanish in the expansion of the log-likelihood ratio, due to the imposed orthogonality condition

$$\text{vec}(B_T)' (\beta'_\perp (\beta_\perp \Psi^{-1} \alpha'_\perp \mu) \otimes \alpha') = \text{vec}(\alpha B \beta'_\perp (\beta_\perp \Psi^{-1} \alpha'_\perp \mu))' = 0.$$

- (iii) For $d_T = 0$ (known cointegration rank), $\alpha'_\perp \mu = 0$, and perturbations m_T such that $\alpha'_\perp m_T = 0$, i.e. none of the components of X_t has a linear trend, we obtain a previous Local Asymptotic Mixed Normality (LAMN) result

by Hodgson (1998b). Without imposing $\alpha'_\perp \mu = 0$, for the same local parameter values such that $\alpha'_\perp m_T = 0$, the sequence of local subexperiments is *not* LAMN; see Remark A.2 in the appendix for further details. \square

For the special case $B_T = 0$, we thus have the traditional Locally Asymptotically Normal (LAN) limiting behavior, i.e. the limit of the local subexperiments associated with $B_T = 0$ is a Gaussian shift experiment; we organize this special case into the following corollary.

Corollary 3.1. *Let Assumptions 1-3 hold, $\vartheta \in \Theta$, and $f \in \mathcal{F}_2$. Then, the sequences of local subexperiments associated with perturbations of the form (3.1)-(3.3) with $B_T = 0$ are asymptotically normal.* \square

Remark 3.2. The above LAN behavior will be used to motivate tests on the cointegration rank. However, as the cointegrating vectors β have to be estimated (imposing the null hypothesis), in order to calculate the (ranks of the) residuals, we are forced to consider experiments with $B_T \neq 0$ as well. As explained before, this leads to non-LAMN behaviour. We will demonstrate in Section 4 that, for our specific testing problem, adaptivity with respect to the perturbation B_T holds, so that optimality results carry over to the non-LAN model of interest.

3.2. Optimal parametric inference on the cointegration rank

We are interested in the problem of testing the null hypothesis that the cointegration rank is r_0 against the alternative value $r_0 + 1$. In the presence of a deterministic linear trend, this is achieved by detecting the existence of an additional cointegrating vector. Locally, there is exactly one such cointegrating vector, proportional to $\beta_\perp \Psi^{-1} \alpha'_\perp \mu$, at rate $T^{3/2}$; see (3.2).¹¹ In terms of the local parameters (3.1)-(3.3), the null hypothesis of interest is thus $H : d = 0$ versus $H' : d \neq 0$, with the parameters m , G , A , b , and B , the scatter matrix Σ , and the radial density f , playing the role of nuisance parameters.

¹¹Our general result on limit experiments (Proposition A.2 in Appendix A) actually can be used, via the perturbation D , as defined in (A.2), at rate T , to handle contiguous alternatives of the form $r = r_0 + k$ with $k > 1$. In this paper, however, we restrict to $k = 1$.

As a preliminary, however, in this section, we restrict our study to the *parametric* subexperiments in which $B = 0$ and (Σ, f) are known; it follows from Corollary 3.1 that these subexperiments are of the LAN type. It is well known that optimal inference in LAN experiments should be based on the so-called *efficient central sequence* (or *score*), the form of which is given in the following proposition.

Proposition 3.2. *Let Assumptions 1-3 hold, $\vartheta \in \Theta$, $f \in \mathcal{F}_2$, with $r = r_0$ and $\alpha'_\perp \mu \neq 0_{p-r}$. Then,*

- (i) *in the Gaussian shift limit experiment considered in Corollary 3.1, the efficient central sequence for d , when the local parameters m , G , A , and b are treated as nuisances, is*

$$\Delta_d^\star = |\beta_\perp \Psi^{-1} \alpha'_\perp \mu|^2 \int_{u=0}^1 \left(u - \frac{1}{2}\right) d[(I_p - P_\alpha) \alpha_\perp]' W_\phi(u), \quad (3.10)$$

where W_ϕ is a Brownian motion with $\text{var } W_\phi(1) = p^{-1} \mathcal{I}_p(f) \Sigma^{-1}$ and $P_\alpha = \alpha (\alpha' \Sigma^{-1} \alpha)^{-1} \alpha' \Sigma^{-1}$ is a (non-orthogonal) projection matrix P_α , with the convention that $P_\alpha = 0_{p \times p}$ (and $\alpha_\perp = I_p$) in case $r_0 = 0$;

- (ii) *for the corresponding local sequence of cointegration experiments, under Assumptions 1-3, a version of that efficient central sequence is*

$$\Delta_d^{(T)\star} = |\beta_\perp \Psi^{-1} \alpha'_\perp \mu|^2 [(I_p - P_\alpha) \alpha_\perp]' \Sigma^{-1/2} S_f^{(T)}, \quad (3.11)$$

with

$$S_f^{(T)} = S_f^{(T)}(\vartheta; \Sigma) := T^{-1/2} \sum_{t=1}^T \left(\frac{t}{T+1} - \frac{1}{2} \right) U_t(\vartheta; \Sigma) \phi_f(\|\epsilon_t(\vartheta)\|_\Sigma). \quad (3.12)$$

□

Exploiting further the classical theory of LAN experiments, the following results are easily obtained; the proof follows along the lines of the more complicated proof of Proposition 4.3 and details are left to the reader.

- (i) The quadratic form

$$Q_f^{(T)}(\vartheta, \Sigma) := \frac{12p}{\mathcal{I}_p(f)} S_f^{(T)}(\vartheta; \Sigma)' \Sigma^{-1/2} \mathcal{Q}_{\alpha, \Sigma} \Sigma^{-1/2} S_f^{(T)}(\vartheta; \Sigma) \quad (3.13)$$

with $Q_f^{(T)}(\vartheta, \Sigma) := (12p/\mathcal{I}_p(f)) \left| S_f^{(T)}(\vartheta; \Sigma) \right|^2$ in case $r_0 = 0$ and

$$\mathcal{Q}_{\alpha, \Sigma} := \left[(I_p - P_\alpha) \alpha_\perp \right] \left[\alpha'_\perp \left(\Sigma^{-1} - \Sigma^{-1} \alpha (\alpha' \Sigma^{-1} \alpha)^{-1} \alpha' \Sigma^{-1} \right) \alpha_\perp \right]^{-1} \left[(I_p - P_\alpha) \alpha_\perp \right]' \quad (3.14)$$

is asymptotically chi-square with $(p - r_0)$ degrees of freedom under the null hypothesis.

(ii) Substituting appropriate (see Assumptions 4 and 5) estimators $(\hat{\vartheta}^{(T)}, \hat{\Sigma}^{(T)})$ for (ϑ, Σ) has no impact, asymptotically, on $Q_f^{(T)}(\vartheta, \Sigma)$.

(iii) The test rejecting the null hypothesis whenever $Q_f^{(T)}(\hat{\vartheta}^{(T)}, \hat{\Sigma}^{(T)})$ exceeds the $(1 - z)$ quantile of the chi-square distribution with $(p - r_0)$ degrees of freedom is asymptotically most stringent at probability level z .

Of course, the procedure described under (iii) is not implementable in the *semiparametric* model of interest, as it strongly depends on f being the actual density; this is why we now introduce its pseudo-Gaussian and rank-based counterparts.

4. Semiparametric inference

This section turns to the semiparametric—unspecified radial density f —version of the testing problem discussed in Section 3.2. In essence, the approach consists in replacing $S_f^{(T)}$ in the quadratic form $Q_f^{(T)}$ given in (3.13) by some computable (that is, not involving the actual, unspecified, density) counterpart. The traditional way of doing this is adopted in Section 4.1, and consists in showing that the tests based on the Gaussian version $Q_\phi^{(T)}$ of (3.13) remain valid, contingent on mild regularity conditions, under non-Gaussian densities, which can be interpreted as a QMLE property. As can be expected, this pseudo-Gaussian test is optimal at Gaussian innovations only; away from the Gaussian, its performances may be poor. The second approach relies on a rank-based version of $Q_f^{(T)}$ and allows for (asymptotically) distribution-free tests reaching optimality also at possibly non-Gaussian reference densities. That approach is developed in Section 4.3, while Section 4.2 explains how ranks naturally come

into the picture in a wide range of semiparametric problems, including those considered here.

Both approaches require consistent pre-estimation, *under the null hypothesis* that $r = r_0$, of the nuisances Σ , μ , Γ , α , and β (at their proper rates of convergence). For the estimation of Σ , we need the following assumption, which is satisfied by all concepts of scatter considered in the literature, among which the empirical covariance matrix, Tyler (1987)'s robust estimator, as well as the R-estimators of Hallin et al. (2006) (when computed from the residuals).

Assumption 4. The sequence of estimators $\hat{\Sigma}^{(T)}$ is such that

- (i) for some $a > 0$, $T^{1/2}(\hat{\Sigma}^{(T)} - a\Sigma)$ is $O_P(1)$ under $P_{\theta; \Sigma, f}^{(T)}$, as $T \rightarrow \infty$;
- (ii) $\hat{\Sigma}^{(T)}$ is a measurable function of the ϵ_t 's, and is invariant under their permutations and reflections with respect to the origin. \square

In the formal analysis of the testing procedures, we will need an assumption of *local asymptotic discreteness* on the estimators of $(\mu, \Gamma, \alpha, \beta)$. This concept is well known for uniform $T^{1/2}$ consistency rates, but needs a refinement in order to handle the mixed T and $T^{3/2}$ rates associated with the cointegrating vectors β . Moreover, the analysis is complicated by the fact that these rates are associated with directions that themselves depend on an unknown parameter, namely $\beta_{\perp} \Psi^{-1} \alpha'_{\perp} \mu$. Therefore, we formulate the definition of local asymptotic discreteness in a somewhat nonstandard form. A sequence of estimators $\hat{\theta}^{(T)} = (\hat{\mu}^{(T)}, \hat{\Gamma}^{(T)}, \hat{\alpha}^{(T)} \hat{\beta}^{(T)'})$ of $\theta = (\mu, \Gamma, \Pi)$ is called *locally asymptotically discrete* if it satisfies the following assumption.¹²

Assumption 5. The estimation errors

$$T^{1/2}(\hat{\mu}^{(T)} - \mu), T^{1/2}(\hat{\Gamma}^{(T)} - \Gamma), T^{1/2}(\hat{\alpha}^{(T)} - \alpha), T(\hat{\beta}^{(T)} - \beta),$$

and

$$T^{3/2}(\hat{\beta}^{(T)} - \beta)' \beta_{\perp} \Psi^{-1} \alpha'_{\perp} \mu$$

¹²As α and β as well as $\hat{\alpha}^{(T)}$ and $\hat{\beta}^{(T)}$ are not uniquely identified, we implicitly impose that it is possible to select versions of these objects such that the assumption holds.

all are $O_P(1)$ under $P_{\vartheta; \Sigma, f}^{(T)}$ as $T \rightarrow \infty$ (rate optimality). Moreover, for any $M > 0$, the number of distinct possible values of these estimation errors in balls of radius M centered at the origin, is bounded as $T \rightarrow \infty$. \square

In standard situations, where all parameters are estimable at rate $T^{1/2}$, any $T^{1/2}$ -consistent estimator can easily be turned into a locally asymptotically discrete one by simply rounding each element to the closest point in a grid of the form $\{kT^{-1/2} : k \in \mathbb{Z}\}$. Indeed, such rounding does not affect $T^{1/2}$ -consistency and leads to the desired discreteness. In the ECM with deterministic linear trend this is, due to the variety of convergence rates, a bit more complicated; see Appendix B for the construction of a locally asymptotically discrete estimator starting from a rate-optimal one. It should be insisted, though, that local asymptotic discreteness is required in formal asymptotic statements, but has little to no practical implications. Certainly, one should not bother to discretize estimators in practice: see Page 125 or 188 of Le Cam and Yang (1990) for a discussion on this point.

4.1. Pseudo-Gaussian testing procedures

Although we throughout avoid Gaussian assumptions and emphasize the semi-parametric nature of the problem, Gaussian procedures—more precisely, the pseudo- or quasi-Gaussian ones—remain a classical benchmark. More, these procedures have the advantage of not requiring the ellipticity assumption on the innovations (but merely a zero mean). However, as we will see, this advantage comes at a significant cost of reduced power.

Particularizing (3.13) to the case of a Gaussian \mathbf{f} , namely, letting $\mathbf{f} = \mathbf{f}_G$ with $\mathbf{f}_G(e) = (2\pi)^{-p/2} |\Sigma|^{-1/2} \exp(-\frac{1}{2} \|e\|_{\Sigma}^2)$, hence $f_G(z) = 2^{1-p/2} \Gamma^{-1}(\frac{p}{2}) \exp(\frac{-z^2}{2})$ and $\phi_f(z) = z$, the efficient central sequence (3.11) takes the form

$$\Delta_{d_{\dagger}}^{(T)\star} = |\beta_{\perp} \Psi^{-1} \alpha'_{\perp} \mu|^2 [(I_p - P_{\alpha}) \alpha_{\perp}]' \Sigma^{-1/2} S_{\dagger}^{(T)}, \text{ with}$$

$$S_{\dagger}^{(T)} := S_{\dagger}^{(T)}(\vartheta, \Sigma) := T^{-1/2} \sum_{t=1}^T \left(\frac{t}{T+1} - \frac{1}{2} \right) \Sigma^{-1/2} \epsilon_t(\vartheta),$$

yielding, for (3.13), the quadratic form (4.1) below.

Proposition 4.1. *Let Assumptions 1-3 hold, $\vartheta \in \Theta$, $f \in \mathcal{F}_2$, with $r = r_0$ and $\alpha'_\perp \mu \neq 0_{p-r}$. Consider the quadratic form*

$$Q_\dagger^{(T)}(\vartheta, \Sigma) := 12S_\dagger^{(T)}(\vartheta, \Sigma)' \Sigma^{-1/2} \mathcal{Q}_{\alpha, \Sigma} \Sigma^{-1/2} S_\dagger^{(T)}(\vartheta, \Sigma), \quad (4.1)$$

with $Q_\dagger^{(T)}(\vartheta, \Sigma) := 12 \left| S_\dagger^{(T)}(\vartheta, \Sigma) \right|^2$ in case $r_0 = 0$, and $\mathcal{Q}_{\alpha, \Sigma}$ defined in (3.14). Then,

- (i) for any $f \in \mathcal{F}_2$, $Q_\dagger^{(T)}(\vartheta, \Sigma)$ under $P_{\vartheta, \Sigma; f}^{(T)}$ has an asymptotic $\chi^2_{p-r_0}$ distribution (this distribution is exact for Gaussian \mathfrak{f});
- (ii) for any $f \in \mathcal{F}_2$, under $P_{\vartheta^{(T)}, \Sigma; f}$ with $\vartheta^{(T)}$ as in (3.1)–(3.3), $Q_\dagger^{(T)}(\vartheta, \Sigma)$ has a limiting non-central $\chi^2_{p-r_0}$ distribution with noncentrality parameter

$$\frac{\mathcal{I}_p^2(f, f_G)}{12p^2} \left| \beta_\perp \Psi^{-1} \alpha'_\perp \mu \right|^4 d' \alpha'_\perp \Sigma^{-1} \mathcal{Q}_{\alpha, \Sigma} \Sigma^{-1} \alpha_\perp d, \quad (4.2)$$

where

$$\mathcal{I}_p(f, g) := \int_0^1 \phi_f(\tilde{F}_p^{-1}(u)) \phi_g(\tilde{G}_p^{-1}(u)) du, \quad (4.3)$$

for $r_0 > 0$, and $(\mathcal{I}_p^2(f, f_G)/12p^2) \left| \beta_\perp \Psi^{-1} \alpha'_\perp \mu \right|^4 \|d\|_\Sigma^2$ in case $r_0 = 0$;

- (iii) the limiting $\chi^2_{p-r_0}$ and non-central $\chi^2_{p-r_0}$ distributions in (i) and (ii) remain asymptotically valid for $Q_\dagger^{(T)}(\hat{\vartheta}^{(T)}, \hat{\Sigma}^{(T)})$, where the estimators $\hat{\Sigma}^{(T)}$ and $\hat{\vartheta}^{(T)}$ satisfy Assumptions 4 and 5, respectively, and the constraint that $r = r_0$;
- (iv) when $\beta_\perp \Psi^{-1} \alpha'_\perp \mu \neq 0$, the (quasi-or pseudo-Gaussian) test rejecting the null hypothesis of a cointegration rank $r = r_0$ (with unspecified elliptical density \mathfrak{f}) whenever $Q_\dagger^{(T)}(\hat{\vartheta}^{(T)}, \hat{\Sigma}^{(T)})$ exceeds the $(1 - z)$ -quantile of a chi-square distribution with $p - r_0$ degrees of freedom is locally and asymptotically z -level most stringent, and reaches parametric efficiency against alternatives of the local form $H' : d \neq 0$ with Gaussian density \mathfrak{f} . \square

The tests described in Part (iv) of Proposition 4.1 are Lagrange multiplier counterparts of Johansen's Gaussian likelihood ratio tests (Johansen 1991 and 1995); both qualify as pseudo-Gaussian tests, as their validity extends to all radial densities f with finite second-order moments. Quite remarkably, the

dependence on f of local powers (the noncentrality parameters (4.2)) is entirely characterized by the scalar *cross-information quantity* $\mathcal{I}_p(f, f_G)$ defined in (4.3). Those cross-information quantities are exactly the same as in the location problems considered in Hallin and Paindaveine (2002a).

Pseudo-Gaussian tests are asymptotically optimal under Gaussian radial densities. The next section proposes rank-based tests leading to a class of tests that can achieve optimality at any radial density (satisfying mild regularity assumptions).

4.2. Optimal rank-based inference

We now turn to rank-based inference. Before providing formal arguments, consider the following intuition. In the semiparametric model, the (radial) density of the innovations ε_t is unknown. As in the univariate case, the multivariate (signed) ranks of those innovations (a precise definition of which is provided below), are *maximally invariant*¹³ with respect to some appropriate groups acting on the observations and generating the various possible radial densities¹⁴. Invariance, in particular, implies distribution-freeness (here, the distribution of the ranks is the same irrespective of the underlying (radial) innovation density); and, all invariant statistics are measurable with respect to a maximal invariant (hence, in the present case, only the rank-based statistics are invariant). Now, in the semiparametric situation where the innovation radial density f is unknown, a statistic conveying, about the parameter of interest, information that does not depend on the nuisances f and Σ has to be invariant with respect to that group \mathcal{G} ; being invariant, it has to be rank-based. Intuition therefore suggests

¹³Recall (see Page 214 of Lehmann and Romano (2005)) that a (measurable) function $\Upsilon(X)$ is invariant under some group of transformations \mathcal{G} acting on X if $\Upsilon(gX) = \Upsilon(X)$ for all $g \in \mathcal{G}$, and that it is maximal invariant if moreover $\Upsilon(Y) = \Upsilon(X)$ implies that $Y = gX$ for some $g \in \mathcal{G}$.

¹⁴that is, for any couple of radial densities f_1, f_2 , there exists a $g \in \mathcal{G}$ such that, if the distribution of X is characterized by radial density f_1 , the distribution of gX is characterized by radial density f_2 .

that ranks do not carry any information about f , and retain that information about the parameter of interest, θ , say, which does not depend on f and Σ . That intuition turns out to be correct, as we now formally show.

More formally, consider the model in which Σ is known and f is unknown, i.e. the semiparametric model

$$\mathcal{P}^{(T)} \left(= \mathcal{P}_{\Sigma}^{(T)} \right) := \left\{ P_{\vartheta; \Sigma, f}^{(T)} \mid \vartheta \in \Theta, f \in \mathcal{F}_2 \right\}.$$

Recall that our problem turns out to be adaptive with respect to Σ , so that efficiency bounds do not depend on whether Σ is known or not. Semiparametric efficiency bounds (in the sense of Bickel et al. (1993)), whether considered at some chosen *reference density* $g \in \mathcal{F}_2$ or uniformly over all densities $g \in \mathcal{F}_2$, provide the relevant optimality concept and characterize the best performances one can hope for in such models.

Rank-based tests in that context are naturally motivated by classical invariance arguments, while remaining compatible with efficiency arguments. Indeed, general results by Hallin and Werker (2003) indicate that if

- (a) the parametric fixed- f submodels $\mathcal{P}_{\Sigma, f}^{(T)} := \{P_{\vartheta; \Sigma, f}^{(T)} \mid \vartheta \in \Theta\}$, $f \in \mathcal{F}$, are Locally Asymptotically Normal (LAN, with respect to ϑ), with *central sequence* $\Delta_{\vartheta; \Sigma, f}^{(T)}$, and
- (b) the nonparametric fixed- ϑ submodels $\mathcal{P}_{\vartheta, \Sigma}^{(T)} := \{P_{\vartheta; \Sigma, f}^{(T)} \mid f \in \mathcal{F}\}$, $\vartheta \in \Theta$, are generated by some group of transformations $\mathcal{G}_{\vartheta; \Sigma}^{(T)}$ acting on $X^{(T)}$, with maximal invariant $M_{\vartheta; \Sigma}^{(T)}$ (typically, a combination of residual ranks and signs),

then, inference procedures reaching, as $T \rightarrow \infty$, the semiparametric efficiency bounds at $P_{\vartheta; \Sigma, g}^{(T)}$ for given $g \in \mathcal{F}$ can be based on the conditional expectation $\underline{\Delta}_{\vartheta; \Sigma, g}^{(T)} := E_{\vartheta; \Sigma, g} \left[\Delta_{\vartheta; \Sigma, g}^{(T)} \mid M_{\vartheta; \Sigma}^{(T)} \right]$ (where $E_{\vartheta; \Sigma, g}$ stands for expectation under $P_{\vartheta; \Sigma, g}^{(T)}$). In other words, $\underline{\Delta}_{\vartheta; \Sigma, g}^{(T)}$ is a version of the semiparametrically efficient, *at ϑ and g* , central sequence for the semiparametric model $\mathcal{P}^{(T)}$. This version of the central sequence, obtained by projecting on on the σ -field generated by the multivariate ranks and signs, precisely formalizes the idea formulated at

the beginning of this section. The fact that $\Delta_{\vartheta;\Sigma,g}^{(T)}$ is measurable with respect to the maximal invariant of a generating group implies that, contrary to the traditional semiparametrically efficient central sequence resulting from tangent space projections, it is distribution-free under $\mathcal{P}_{\vartheta;\Sigma}^{(T)}$. Inference based on $\Delta_{\vartheta;\Sigma,g}^{(T)}$ thus remains valid irrespective of the actual density f , and g here plays the role of a *reference density*, to be chosen by the researcher. Still without affecting validity, the choice of that reference density also can be data-driven—for instance, an estimator $\hat{f}^{(T)}$ of f can be substituted for g ; we refer to Hallin and Werker (2003) for details.

In the present context, the maximal invariant $M_{\vartheta;\Sigma}^{(T)}$ happens to be the T -tuple of multivariate signs $U_t(\vartheta;\Sigma)$ defined in (3.5), along with the ranks $R_t^{(T)}(\vartheta;\Sigma)$ of the norms $\|\epsilon_t(\vartheta)\|_{\Sigma}$ of the residuals $\epsilon_t(\vartheta)$ (see (3.4)). To see this formally, fixing ϑ and Σ , consider the group $\mathcal{G}_{\vartheta;\Sigma}^{(T),\circ}$ of transformations $\mathcal{G}_{\vartheta;\Sigma}^{\mathbf{m}}$ of \mathbb{R}^{pT} , with \circ denoting composition of transformations, and indexed by $\mathbf{m} \in \mathfrak{M}$, with

$$\mathfrak{M} := \{\mathbf{m} : \mathbb{R}^+ \rightarrow \mathbb{R}^+ \mid \mathbf{m}(0) = 0, \lim_{z \rightarrow \infty} \mathbf{m}(z) = \infty, \mathbf{m} \text{ monotone increasing}\} \cap C(\mathbb{R}_+),$$

where $\mathcal{G}_{\vartheta;\Sigma}^{\mathbf{m}}$ is mapping the series $X^{(T)}$ with p -dimensional observations X_t onto the transformed series $X^{\mathbf{m};(T)} := \mathcal{G}_{\vartheta;\Sigma}^{\mathbf{m}}(X^{(T)})$ with p -dimensional observations

$$X_t^{\mathbf{m}} := (I_p + \Pi)X_{t-1}^{\mathbf{m}} + \sum_{j=1}^{k-1} \Gamma_j \Delta X_{t-j}^{\mathbf{m}} + \mu + \mathbf{m}(\|\epsilon_t(\vartheta)\|_{\Sigma}) \Sigma^{1/2} U_t(\vartheta;\Sigma), \quad t = 1, \dots, T$$

(a recursive definition, with the k deterministic starting values remaining unchanged). Letting $\mathbf{m} = \tilde{G}_p^{-1} \circ \tilde{F}_p$, it is easy to see that the joint distribution of $X^{(T)}$ is $P_{\vartheta;\Sigma,f}^{(T)}$ if and only if the joint distribution of $\mathcal{G}_{\vartheta;\Sigma}^{\mathbf{m}}(X^{(T)})$ is $P_{\vartheta;\Sigma,g}^{(T)}$, so that $\mathcal{G}_{\vartheta;\Sigma}^{(T),\circ}$ is a generating group for $\mathcal{P}_{\vartheta;\Sigma}^{(T)}$, with maximal¹⁵ invariant the multivariate signs $U_t(\vartheta;\Sigma)$ and ranks $R_t^{(T)}(\vartheta;\Sigma)$, as announced.

¹⁵For the maximal invariance of $(U_1(\vartheta;\Sigma), \dots, U_T(\vartheta;\Sigma); R_1^{(T)}(\vartheta;\Sigma), \dots, R_T^{(T)}(\vartheta;\Sigma))$, it is required to show that two points x and y in the observation space belong to the same orbit of $\mathcal{G}_{\vartheta;\Sigma}^{(T)}$ if and only if they yield the same U_i 's and the same R_i 's. That belonging to the same orbit implies having the same U_i 's and the same R_i 's is straightforward. The converse (two points x and y sharing the same U_i 's and the same R_i 's belong to the same orbit) is shown to hold by exhibiting a transformation that belongs to the group and maps x onto y . Such a transformation can be constructed exactly as in the traditional one-dimensional signed-rank case—see Lehmann and Romano (2005, p.242).

Those multivariate signed ranks have been successfully used in a series of papers by Hallin and Paindaveine (2002a, 2002b, 2005b, 2006), Hallin et al. (2006) and Hallin et al. (2010, 2012) for various problems (hypothesis testing and point estimation) in multivariate analysis and multivariate (stationary) time series. All those papers consider models for which (a) is satisfied. However, for the cointegration model, (a) does *not* hold. Invoking adaptivity arguments, we show that it is nevertheless possible to develop semiparametrically (and even parametrically) efficient tests using signed ranks.

4.3. A new rank-based test

Optimal inference in the parametric submodels of Section 3.2 was based on statistics of the form $S_f^{(T)}$ (see (3.12)) which, via ϕ_f , depend on the unknown radial density, hence cannot be computed, in the semiparametric model, from the observations. For a given reference density g , we propose to replace $S_f^{(T)}$ with

$$\mathcal{S}_g^{(T)}(\vartheta; \Sigma) := T^{-1/2} \sum_{t=1}^T \left(\frac{t}{T+1} - \frac{1}{2} \right) U_t(\vartheta; \Sigma) \phi_g \left(\tilde{G}_p^{-1} \left(\frac{R_t^{(T)}(\vartheta; \Sigma)}{T+1} \right) \right). \quad (4.4)$$

As $\sum_{t=1}^T (t/(T+1) - 1/2) = 0$, $\mathcal{S}_g^{(T)}(\vartheta; \Sigma)$ is centered under any $P_{\vartheta; \Sigma, f}$. A straightforward application of traditional results on linear rank statistics (see, e.g., Hájek and Sidák (1967), Section V.1.6) now shows that the so-called *approximate score* form (4.4) of $\mathcal{S}_g^{(T)}$ is asymptotically equivalent under $P_{\vartheta; \Sigma, g}$ to the *exact score* form

$$T^{-1/2} \sum_{t=1}^T \left(\frac{t}{T+1} - \frac{1}{2} \right) E_{\vartheta; \Sigma, g} \left[U_t(\vartheta; \Sigma) \phi_g (\|\epsilon_t(\vartheta)\|_\Sigma) \mid U_t(\vartheta; \Sigma), R_t^{(T)}(\vartheta; \Sigma) \right],$$

so that, in line with the ideas developed in Section 4.2, the difference between $\mathcal{S}_g^{(T)}(\vartheta; \Sigma)$ and $E_{\vartheta; \Sigma, g}[\mathcal{S}_g^{(T)} | M_{\vartheta; \Sigma}^{(T)}]$ is $o_P(1)$ under any $P_{\vartheta; \Sigma, f}$.

Proposition 4.2 on the asymptotic behavior of rank-based statistics requires the following classical assumption on the reference density g (see, for example, Hájek and Sidák (1967), p.164).

Assumption 6. The (radial) reference density g satisfies Assumption 1, and is such that ϕ_g is the difference of two continuous and monotone increasing functions. \square

We now can state the main results on the asymptotic behavior of $\underline{S}_g^{(T)}(\vartheta; \Sigma)$.

Proposition 4.2. *Let Assumptions 1-6 hold, $\vartheta \in \Theta$ and $f \in \mathcal{F}_2$. Consider a bounded sequence of perturbations*

$$h'_T := (m'_T, (\text{vec } G_T)', (\text{vec } A_T)', b'_T, (\text{vec } B_T)', d'_T), \quad (4.5)$$

with $\beta_\perp B'_T \perp (\beta_\perp \Psi_{\Gamma, \Pi}^{-1} \alpha'_\perp \mu)$, defining, as in (3.1)-(3.3), a local parameter sequence $\vartheta^{(T)}$. Then, as $T \rightarrow \infty$,

(i) (asymptotic representation) under $P_{\vartheta^{(T)}; \Sigma, f}^{(T)}$,

$$\underline{S}_g^{(T)}(\vartheta^{(T)}; \hat{\Sigma}^{(T)}) = \underline{S}_g^{(T)}(\vartheta^{(T)}; \Sigma) + o_P(1), \quad (4.6)$$

where

$$\underline{S}_g^{(T)}(\vartheta^{(T)}; \Sigma) := T^{-1/2} \sum_{t=1}^T \left(\frac{t}{T+1} - \frac{1}{2} \right) U_t(\vartheta^{(T)}; \Sigma) \phi_g \left(\tilde{G}_p^{-1}(\tilde{F}_p(\|\epsilon_t(\vartheta^{(T)})\|_\Sigma)) \right);$$

(ii) (asymptotic normality under the null) under $P_{\vartheta; \Sigma, f}^{(T)}$, $\underline{S}_g^{(T)}(\vartheta; \hat{\Sigma}^{(T)})$ is asymptotically normal, with mean zero and variance $\frac{1}{12p} \mathcal{I}_p(g) I_p$;

(iii) (asymptotic normality under alternatives) if also $B_T = 0$, $\underline{S}_g^{(T)}(\vartheta; \hat{\Sigma}^{(T)})$ is asymptotically normal under $P_{\vartheta^{(T)}; \Sigma, f}^{(T)}$, with mean

$$\frac{1}{12p} \mathcal{I}_p(f, g) |\beta_\perp \Psi^{-1} \alpha'_\perp \mu|^2 \Sigma^{-1/2} (\alpha b + \alpha_\perp d) \quad (4.7)$$

and variance $\frac{1}{12p} \mathcal{I}_p(g) I_p$;

(iv) (asymptotic linearity) moreover, under $P_{\vartheta; \Sigma, f}^{(T)}$,

$$\underline{S}_g^{(T)}(\vartheta^{(T)}; \hat{\Sigma}^{(T)}) - \underline{S}_g^{(T)}(\vartheta; \hat{\Sigma}^{(T)}) = -\frac{\mathcal{I}_p(f, g)}{pT^{1/2}} \sum_{t=1}^T \left(\frac{t}{T+1} - \frac{1}{2} \right) v_{Tt} + o_P(1), \quad (4.8)$$

with $v_{Tt} = (v_{Tt,1}, \dots, v_{Tt,p})'$, where $(e_j$ stands for the j th unit vector in the canonical basis of \mathbb{R}^p)

$$v_{Tt,j} := h'_T \begin{pmatrix} Z_{Tt}^{(1)}(\vartheta) \otimes \Sigma^{-1/2} e_j \\ Z_{Tt}^{(2)}(\vartheta) \otimes \alpha' \Sigma^{-1/2} e_j \\ Z_{Tt}^{(2)}(\vartheta) \otimes \alpha'_\perp \Sigma^{-1/2} e_j \end{pmatrix}, \quad j = 1, \dots, p,$$

and $Z_{Tt}^{(1)}, Z_{Tt}^{(2)}$ are defined in (3.8) and (3.9);

(v) the asymptotic linearity property (4.8) remains valid if $\vartheta^{(T)}$ is replaced with a locally asymptotically discrete (in the sense of Assumption 5) random sequence $\hat{\vartheta}^{(T)}$. \square

Part (iii) of Proposition 4.2 assumes $B_T = 0$. The reason is that nonzero B_T perturbations lead to non-LAN behavior. As a result, the limiting distributions of $\underline{S}_g^{(T)}(\vartheta^{(T)}; \hat{\Sigma}^{(T)})$ are no longer of the Gaussian shift type. While exact calculations (under $B_T \neq 0$) are possible, they do not necessarily provide much insight and, therefore, are omitted. Note, however, that the estimators used in the computation of aligned ranks, of course, may deviate from the actual parameters in the $B_T \neq 0$ direction: accordingly, the assumption that $B_T = 0$ is not made in parts (iv) and (v) of the proposition. This is the reason we had to consider these non-LAN directions in Proposition 3.1.

From Proposition 4.2, we see that the rank-based statistic $\underline{S}_g^{(T)}$ is well-behaved even under misspecified reference density. Under correctly specified reference density ($g = f$), $\underline{S}_g^{(T)}$ is asymptotically equivalent to the parametric statistic $\underline{S}_f^{(T)} = S_f^{(T)}$, hence enjoys the same properties as the latter, which justifies its interpretation as a rank-based version of $S_f^{(T)}$. Note that the asymptotic representation result in Part (i) shows that estimating the scatter matrix Σ has no asymptotic impact on $\underline{S}_g^{(T)}(\vartheta^{(T)}; \hat{\Sigma}^{(T)})$, neither under the null nor under contiguous alternatives.

Proposition 4.2 then can be used as a basis for the construction of locally and asymptotically optimal rank-based tests.

Proposition 4.3. *Let Assumptions 1-6 hold, $\vartheta \in \Theta$ be such that $r = r_0$, and $f \in \mathcal{F}_2$. Consider the rank-based statistic $\underline{S}_g^{(T)}(\vartheta; \Sigma)$ in (4.4) and the quadratic form*

$$\underline{Q}_g^{(T)}(\vartheta, \Sigma) := \frac{12p}{\mathcal{I}_p(g)} \underline{S}_g^{(T)}(\vartheta; \Sigma)' \Sigma^{-1/2} \mathcal{Q}_{\alpha, \Sigma} \Sigma^{-1/2} \underline{S}_g^{(T)}(\vartheta; \Sigma), \quad (4.9)$$

with $\underline{Q}_g^{(T)}(\vartheta, \Sigma) := (12p/\mathcal{I}_p(g)) \left| \underline{S}_g^{(T)}(\vartheta; \Sigma) \right|^2$ in case $r_0 = 0$, and $\mathcal{Q}_{\alpha, \Sigma}$ defined in (3.14) Then,

(i) $\underline{Q}_g^{(T)}(\vartheta, \Sigma)$ under $P_{\vartheta; \Sigma, f}^{(T)}$ has a limiting $\chi_{p-r_0}^2$ distribution;

(ii) under $P_{\vartheta^{(T)}; \Sigma; f}$, with $\vartheta^{(T)}$ as in (3.1)–(3.3), $Q_g^{(T)}(\vartheta, \Sigma)$ has a limiting non-central $\chi_{p-r_0}^2$ distribution with noncentrality parameter

$$\frac{\mathcal{I}_p(f, g)^2}{12p\mathcal{I}_p(g)} |\beta_{\perp} \Psi^{-1} \alpha'_{\perp} \mu|^4 d' \alpha'_{\perp} \Sigma^{-1} \mathcal{Q}_{\alpha, \Sigma} \Sigma^{-1} \alpha_{\perp} d \quad (4.10)$$

for $r_0 > 0$, and $(\mathcal{I}_p(f, g)^2/12p\mathcal{I}_p(g)) |\beta_{\perp} \Psi^{-1} \alpha'_{\perp} \mu|^4 \|d\|_{\Sigma}^2$ in case $r_0 = 0$;

(iii) those limiting distributions remain valid if the statistic $Q_g^{(T)}(\vartheta, \Sigma)$ is computed on the basis of estimators $\hat{\Sigma}^{(T)}$ and $\hat{\vartheta}^{(T)}$ satisfying Assumptions 4 and 5, respectively, and the constraint that $r = r_0$;

(iv) for $\beta_{\perp} \Psi^{-1} \alpha'_{\perp} \mu \neq 0$ and $f = g$, the test rejecting the null hypothesis of a cointegration rank $r = r_0$ (with unspecified elliptical density f) whenever $Q_g^{(T)}(\hat{\vartheta}^{(T)}, \hat{\Sigma}^{(T)})$ exceeds the $(1 - z)$ -quantile of a chi-square distribution with $(p - r_0)$ degrees of freedom is locally and asymptotically z -level most stringent against alternatives of the local form $H' : d \neq 0$ with radial density g . \square

Note that, if reduced-rank regression methods are used, as in Johansen (1995), to estimate ϑ under the null, the rank-based test statistics $Q_g^{(T)}(\hat{\vartheta}^{(T)}, \hat{\Sigma}^{(T)})$ are invariant with respect to non-singular linear transformations of the data.

The rank-based test statistic $Q_g^{(T)}(\vartheta, \Sigma)$ in Proposition 4.3 has several quite desirable properties. First of all, due to its dependence on the innovation ranks only, it is exactly distribution-free under the null. In particular, note that any scalar factor in the innovation scatter matrix Σ would cancel out of the statistic. Also, the statistic does not depend on the versions chosen for α and β or α_{\perp} and β_{\perp} . The resulting test thus has a constant rejection probability over the null hypothesis. Those results carry over to the asymptotic size when considering the aligned-rank version $Q_g^{(T)}(\hat{\vartheta}^{(T)}, \hat{\Sigma}^{(T)})$. And, contrary to classical semiparametric methods (Bickel et al., 1993), the actual radial density f needs not be estimated for this.

The local and asymptotic power of the test, determined by the noncentrality parameter (4.10), does depend on the actual underlying density f , the scatter

matrix Σ , the drift $\beta_{\perp}\Psi^{-1}\alpha'_{\perp}\mu$, and the column spaces of α and α_{\perp} . Again, this asymptotic power result is not affected by the estimation of ϑ and Σ (under Assumptions 4 and 5). The power of the test gets larger as the reference density g gets closer, as measured by $\mathcal{I}_p(f, g)$, to the actual density f . Also, the power increases with larger values of $|\beta_{\perp}\Psi^{-1}\alpha'_{\perp}\mu|$. For $\alpha'_{\perp}\mu = 0_{p-r}$, the test has asymptotically no power at the rate $T^{3/2}$ in the local alternatives we consider—but note that no test ever would. However, it may very well have power against alternatives at rate T .

Part (iv) of Proposition 4.3 asserts that, for well-chosen reference density g , the rank-based test achieves the parametric efficiency bound in case $\alpha'_{\perp}\mu \neq 0_{p-r}$. In that case, the limiting experiment (still, with $B = 0$) is LAN, so that the concept of efficiency is well defined. The notion that the power of the rank test increases as g gets closer to f suggests the use of a pre-estimated density \hat{f} instead of the fixed reference density g . This idea has been pursued in other contexts, see, e.g., Hallin and Werker (2003), and is equally applicable in the present setting. Incidentally, this shows, as was to be expected, that the inference problem is also adaptive with respect to f . Other data-driven selections of the reference density are also possible, provided that they are based on the order statistic of the residual norms $\|\epsilon_t(\hat{\vartheta}^{(T)}, \hat{\Sigma}^{(T)})\|_{\hat{\Sigma}}$: see, e.g., Hallin and Mehta (2015).

4.4. Nonadmissibility of quasi-Gaussian procedures

It is worth stressing that the Chernoff-Savage property established in Proposition 6 of Hallin and Paindaveine (2002a) also holds here for the normal-score or van der Waerden version of the tests described in Proposition 4.2: the noncentrality parameters of the non-null asymptotic distributions indeed involve the same cross-information quantities. The van der Waerden test here relies on the quadratic form $Q_{\phi}^{(T)}(\hat{\vartheta}^{(T)}, \hat{\Sigma}^{(T)})$ based on $S_{\phi}^{(T)}(\hat{\vartheta}^{(T)}, \hat{\Sigma}^{(T)})$, where

$$S_{\phi}^{(T)}(\vartheta; \Sigma) := T^{-1/2} \sum_{t=1}^T \left(\frac{t}{T+1} - \frac{1}{2} \right) U_t(\vartheta; \Sigma) \left(F_{\chi_p^2}^{-1} \left(\frac{R_t^{(T)}(\vartheta; \Sigma)}{T+1} \right) \right)^{1/2} \quad (4.11)$$

(where $F_{\chi_p^2}$ stands for the chi-square distribution function with p degrees of freedom). This means that asymptotic relative efficiencies under radial density f (ARE_f), with respect to the pseudo- or quasi-Gaussian methods (based on Proposition 4.1), of the van der Waerden rank-based tests (based on (4.11)), are always larger than or equal to one, irrespective of the actual unknown radial density f . Equality is achieved in the Gaussian case only, i.e. for Gaussian f .

This remarkable fact implies that the pseudo-Gaussian tests of Proposition 4.1 are not admissible: one is always strictly better off using van der Waerden rank-based tests rather than the quasi-Gaussian ones, except of course under Gaussian conditions, where both methods, being optimal, perform equally well.

5. Monte Carlo study

This section reports the results of a small Monte Carlo study corroborating our asymptotic analysis and assessing the finite-sample performances of the pseudo-Gaussian and the rank-based tests proposed in Propositions 4.1 and 4.3. We compare their performances to those of Johansen's LR-based tests, namely the *maxeig* (maximum eigenvalue) and *trace* tests. These LR-based tests are often considered in empirical work and, as documented in Hubrich et al. (2001), have comparatively good performances. We also provide these simulation results to show that, at least for the cases considered, the asymptotic results in our paper indeed provide an adequate approximation of the finite-sample distributions. The study is implemented in MATLAB 7.14 and the code is available upon request. The residuals and the estimator of α were obtained via reduced-rank regression.¹⁶ The pseudo-Gaussian tests use the empirical covariance matrix of residuals, and the rank-based tests Tyler's estimator of scatter as estimators

¹⁶We use MATLAB's Econometrics Toolbox to compute the reduced-rank regression estimators as well as the *maxeig* and *trace* statistics.

of Σ .¹⁷ Throughout, the probability level is 5%. The pseudo-Gaussian and rank-based tests are based on the critical values implied by their asymptotic χ^2 -distributions; for the *maxeig* and *trace* tests, we rely on the Matlab tabulated values of MacKinnon et al. (1999).

The data-generating processes (DGPs) considered below are inspired by Hubrich et al. (2001) and Toda (1994); they are described by

$$\Delta X_t = \begin{pmatrix} 0_{(p-1) \times 1} \\ 1 \end{pmatrix} + \begin{pmatrix} \phi I_r & 0_{r \times (p-r)} \\ 0_{(p-r) \times r} & 0_{(p-r) \times (p-r)} \end{pmatrix} X_{t-1} + \varepsilon_t, \quad (5.1)$$

where ϕ is a scalar. Note that $\alpha' = (\phi I_r \quad 0_{r \times (p-r)})$ and $\beta' = (I_r \quad 0_{r \times (p-r)})$; for $\phi \neq 0$, the cointegration rank of $\{X_t\}$ is r . These DGPs can be considered as “canonical forms”, since many interesting models can be obtained by means of nonsingular linear transformations of the X_t ’s; see Hubrich et al. (2001) and Toda (1994) for details. Restricting to such canonical forms is legitimate because the pseudo-Gaussian and rank-based tests we consider, as well as the LR-based *maxeig* and *trace* tests, all are invariant with respect to such transformations.

In the simulation study, we considered $p \in \{2, 3, 5\}$ and $\phi = -0.3$. As innovation radial densities f , we chose Gaussian ϕ , t_3 and t_{10} densities, with covariance matrices $\Sigma_p \in \{I_p, \Sigma_{p,c}\}$, where

$$\Sigma_{2,c} = \begin{pmatrix} 1 & 0.8 \\ & 1 \end{pmatrix}, \quad \Sigma_{3,c} = \begin{pmatrix} 1 & 0.4 & 0.8 \\ & 1 & 0 \\ & & 1 \end{pmatrix}, \quad \text{and} \quad \Sigma_{5,c} = \begin{pmatrix} 1 & 0.4 & 0.8 & 0 & 0 \\ & 1 & 0 & 0 & 0 \\ & & 1 & 0 & 0 \\ & & & 1 & 0 \\ & & & & 1 \end{pmatrix}.$$

The choices of $p = 2$ and 3 are taken from Hubrich et al. (2001) (who did not consider $p = 5$). Still in line with these authors, we used initial values $X_0 = 0_{p \times 1}$, and a “warming up” period of 50 observations (the analysis thus is based on X_{51}, \dots, X_{50+T}). Sample sizes are $T \in \{100, 250, 500\}$. In order to

¹⁷The iterative scheme of Randles (2000, p.1267) is used to compute Tyler’s estimator of scatter (using the Frobenius norm and 10^{-10} as tolerance).

save space, we only show here the results for $p = 5$ and $\Sigma_5 = \Sigma_{5,c}$; the results for $p = 2, 3$ and $p = 5$ with $\Sigma_5 = I_5$ are presented in Appendix D.

To illustrate the quality of the χ^2 approximation of the null distributions of our test statistics, Figure 5.1 presents a scaled histogram of simulated values of the rank-based test statistics $\tilde{Q}_g^{(T)}$, $g \in \{\phi, t_3, t_{10}\}$, under (5.1) for $p = 5$, $r_0 = 0$, $\Sigma_5 = I_5$, and $T = 100$, along with a graph of their limiting χ_5^2 density.

Figure 5.1: Simulated (25,000 replications) finite-sample ($T = 100$) distributions of the rank-based test statistics (4.9), $g \in \{\phi, t_3, t_{10}\}$, for $p = 5$, $r_0 = 0$, $\Sigma_5 = I_5$, compared to their limiting χ_5^2 distribution.

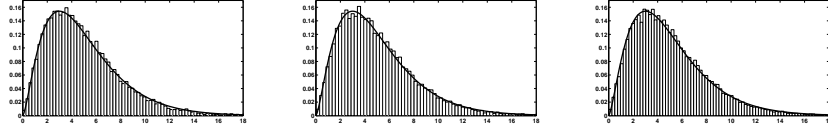


Table 5.1 reports the size of the pseudo-Gaussian test $Q_{\dagger}^{(T)}$, the sizes of the rank-based tests $\tilde{Q}_g^{(T)}$, $g \in \{\phi, t_3, t_{10}\}$, and those of the *maxeig* and *trace* tests under (5.1) for $p = 5$, $r_0 \in \{0, \dots, p-1\}$ and $\Sigma_5 = \Sigma_{5,c}$. The sizes for $\Sigma_5 = I_5$ are shown in Table D.3 in Appendix D. Note that, as was to be expected, the sizes of the rank-based tests are more stable than those of the *maxeig* and *trace* tests. For smaller samples ($T = 100$), the rank-based tests are somewhat undersized for larger r_0 , while the *maxeig* and *trace* ones are slightly oversized for small r_0 . For larger values of T , all sizes are close to the nominal 5% level.

To assess finite-sample powers, we considered alternatives of the form

$$\Pi_h^{(T)} = \Pi + T^{-3/2} \alpha_{\perp} d(\beta_{\perp} \Psi^{-1} \alpha'_{\perp} \mu)', \quad (5.2)$$

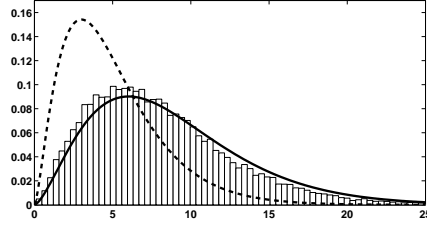
with $d = -h 1_{(p-r) \times 1}$ in (3.2) and the version $\begin{pmatrix} 0_{(p-r) \times r} & I_{(p-r) \times (p-r)} \end{pmatrix}$ of α'_{\perp} . From Proposition 4.3, we know that, under these local alternatives, the rank test statistics are asymptotically non-central χ^2 . As an illustration of this weak convergence result, Figure 5.2 compares the finite-sample distributions of the rank-based test $\tilde{Q}_{\phi}^{(T)}$ under (5.2) with $h = 3$ for $p = 5$, $r_0 = 0$, $\mathbf{f} = \mathcal{N}(0, I_5)$, and $T = 500$, to the corresponding non-central χ^2 distributions. For completeness, we also plot the limiting null distribution once more. Again, we conclude

Table 5.1: Simulated sizes (25,000 replications) of the *maxeig* test, *trace* test, $Q_{\dagger}^{(T)}$ and the rank-based tests (4.9), $g \in \{\phi, t_3, t_{10}\}$, under (5.1) for $p = 5$, $r_0 \in \{0, \dots, 4\}$, $\phi = -0.3$, $\Sigma = \Sigma_{5,c}$, and $f \in \{\phi, t_3, t_{10}\}$. For $r_0 = p-1 = 4$, *maxeig* and *trace* coincide.

Tests	Sample sizes and innovation densities								
	$T = 100$			$T = 250$			$T = 500$		
	$f = \phi$	$f = t_3$	$f = t_{10}$	$f = \phi$	$f = t_3$	$f = t_{10}$	$f = \phi$	$f = t_3$	$f = t_{10}$
$r_0 = 0$									
<i>maxeig</i>	0.064	0.083	0.065	0.053	0.065	0.055	0.049	0.061	0.054
<i>trace</i>	0.071	0.082	0.072	0.058	0.063	0.058	0.054	0.059	0.055
$Q_{\dagger}^{(T)}$	0.044	0.038	0.043	0.047	0.047	0.047	0.048	0.047	0.048
$\tilde{Q}_{\phi}^{(T)}$	0.044	0.043	0.043	0.047	0.049	0.046	0.047	0.048	0.048
$\tilde{Q}_{t_3}^{(T)}$	0.045	0.042	0.046	0.050	0.047	0.047	0.048	0.048	0.049
$\tilde{Q}_{t_{10}}^{(T)}$	0.047	0.045	0.045	0.049	0.048	0.047	0.047	0.047	0.049
$r_0 = 1$									
<i>maxeig</i>	0.072	0.083	0.077	0.057	0.066	0.062	0.054	0.061	0.055
<i>trace</i>	0.082	0.091	0.084	0.061	0.070	0.065	0.054	0.060	0.057
$Q_{\dagger}^{(T)}$	0.035	0.033	0.033	0.042	0.041	0.043	0.046	0.046	0.046
$\tilde{Q}_{\phi}^{(T)}$	0.036	0.041	0.034	0.042	0.046	0.044	0.044	0.048	0.047
$\tilde{Q}_{t_3}^{(T)}$	0.039	0.049	0.040	0.045	0.049	0.045	0.047	0.050	0.047
$\tilde{Q}_{t_{10}}^{(T)}$	0.037	0.045	0.037	0.043	0.049	0.045	0.045	0.049	0.048
$r_0 = 2$									
<i>maxeig</i>	0.029	0.040	0.031	0.053	0.060	0.054	0.051	0.056	0.052
<i>trace</i>	0.036	0.047	0.039	0.053	0.061	0.056	0.052	0.054	0.053
$Q_{\dagger}^{(T)}$	0.022	0.021	0.022	0.038	0.039	0.041	0.044	0.044	0.045
$\tilde{Q}_{\phi}^{(T)}$	0.021	0.028	0.021	0.038	0.044	0.040	0.043	0.047	0.044
$\tilde{Q}_{t_3}^{(T)}$	0.030	0.044	0.034	0.041	0.051	0.043	0.045	0.051	0.046
$\tilde{Q}_{t_{10}}^{(T)}$	0.024	0.034	0.026	0.039	0.048	0.043	0.043	0.050	0.045
$r_0 = 3$									
<i>maxeig</i>	0.020	0.024	0.019	0.048	0.054	0.047	0.050	0.053	0.049
<i>trace</i>	0.021	0.027	0.021	0.049	0.054	0.048	0.049	0.054	0.049
$Q_{\dagger}^{(T)}$	0.017	0.018	0.018	0.039	0.038	0.038	0.042	0.045	0.046
$\tilde{Q}_{\phi}^{(T)}$	0.016	0.022	0.017	0.038	0.041	0.039	0.042	0.046	0.045
$\tilde{Q}_{t_3}^{(T)}$	0.025	0.036	0.027	0.041	0.048	0.042	0.046	0.049	0.045
$\tilde{Q}_{t_{10}}^{(T)}$	0.019	0.026	0.020	0.039	0.043	0.039	0.044	0.047	0.045
$r_0 = 4$									
<i>maxeig</i>	0.040	0.047	0.038	0.044	0.048	0.045	0.046	0.049	0.050
$Q_{\dagger}^{(T)}$	0.026	0.025	0.024	0.040	0.040	0.040	0.044	0.046	0.047
$\tilde{Q}_{\phi}^{(T)}$	0.024	0.026	0.023	0.039	0.042	0.039	0.043	0.047	0.047
$\tilde{Q}_{t_3}^{(T)}$	0.031	0.039	0.032	0.041	0.046	0.044	0.045	0.048	0.048
$\tilde{Q}_{t_{10}}^{(T)}$	0.026	0.036	0.026	0.038	0.044	0.042	0.045	0.048	0.047

that our asymptotic analysis provides a close approximation to these finite-sample performances.

Figure 5.2: Simulated (25,000 replications) finite-sample ($T = 500$) distribution of the van der Waerden test statistic (based on (4.11)) under alternative $\Pi_3^{(T)}$ (5.2) for $p = 5$, $r_0 = 0$ and $\mathbf{f} = \mathcal{N}(0, I_5)$, compared to its χ_5^2 limiting null distribution (dashed) and its non-central χ_5^2 (non-centrality parameter $15/4$) limiting distribution under the alternative (solid).



Figures 5.3 and 5.4 present the power curves of the various tests for $p = 5$ and $r_0 = 0, \dots, 4$. Each figure shows the power, as a function of h in (5.2), for the sample sizes $T \in \{100, 250, 500\}$, $\Sigma = \Sigma_{5_c}$, and innovation radial densities $f \in \{\phi, t_3, t_{10}\}$. The results for the case $\Sigma_5 = I_5$ are presented in Figures D.5 and D.6 in Appendix D. As mentioned before, that appendix also contains the simulation results for $p = 2, 3$.

First of all, we note that the simulation results corroborate the asymptotic result that the power of the rank-based tests is maximal in case $f = g$. The figures also empirically confirm the Chernoff-Savage result of Section 4.4 that the van der Waerden rank-based test dominates its pseudo-Gaussian counterpart, except for Gaussian innovations for which they are equivalent. Those simulation results once more motivate the interest in rank-based tests. If we compare the performances of the pseudo-Gaussian test and the rank-based tests to those of the *maxeig* and *trace* tests, we can make the following observations. For $T = 250, 500$ the pseudo-Gaussian and rank-based tests yield substantial improvements over the *maxeig* and *trace* tests, while for $T = 100$ there is no

uniform conclusion: the *maxeig* and *trace* tests tend to perform better for large values of h . For $r_0 = p - 1$, the pseudo-Gaussian test and rank-based tests do not perform very well, in terms of power, for small sample sizes. Summarizing, we conclude that no uniformly valid conclusions are possible and that the rank-based tests, which are easy to compute, nicely complement the *maxeig* and *trace* tests, and thus constitute a useful addition to the statistical toolkit.

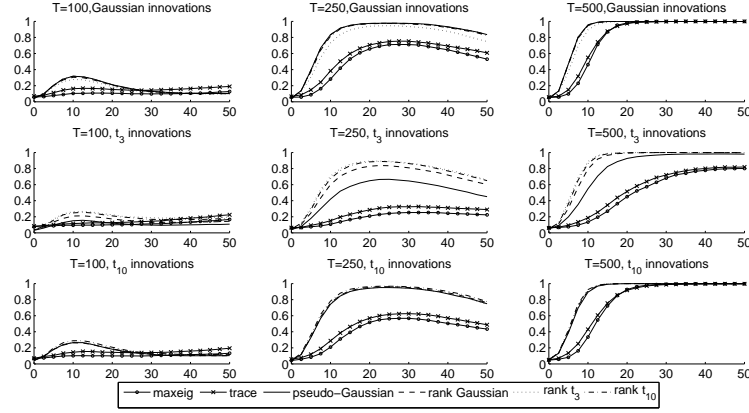
6. Discussion and Conclusion

We analyze optimal testing of the cointegration rank in Error Correction Models. We show that, in applications where linear trends are to be expected, standard Locally Asymptotically Normal behavior can be exploited. We propose rank-based versions of the optimal tests to exploit their advantages, in particular, their very stable size and their QMLE property that extends beyond Gaussian reference densities.

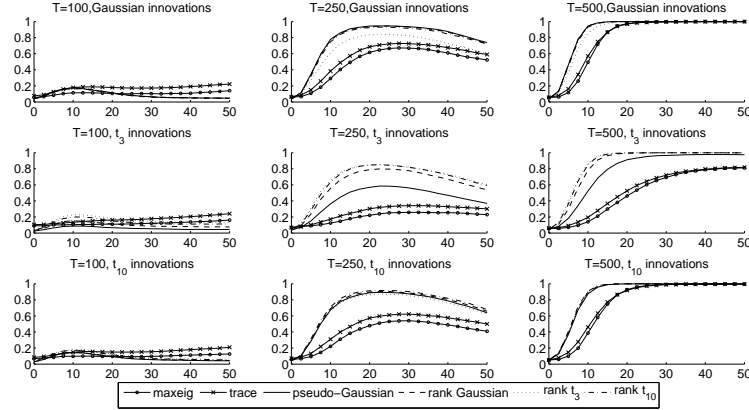
Our full characterization of the limit experiments arising in the Error Correction Model opens interesting avenues for future research. In particular, with respect to the cointegrating vectors, the model is only Locally Asymptotically Mixed Normal in the absence of a linear time trend. Optimality issues for those experiments remain largely unaddressed.

It may be informative to compare our contribution to Boswijk et al. (2015), which improves upon Gaussian likelihood ratio tests by a more careful description of the relevant alternatives. Formally, their alternatives are specified as cones instead of linear subspaces, much like a univariate test for a unit root generally uses the one-sided alternative that the autocorrelation parameter is strictly smaller than one. Unlike ours, that paper uses Gaussian likelihoods; on the other hand, we do not consider the cone-shaped alternatives considered there. Both approaches are complementary and, thus, could possibly be combined.

Figure 5.3: Simulated (2,500 replications) finite-sample powers of the *maxeig* test, the *trace* test, the pseudo-Gaussian test and the rank-based tests (4.9) with reference densities $a \in \{\phi, t_3, t_{10}\}$

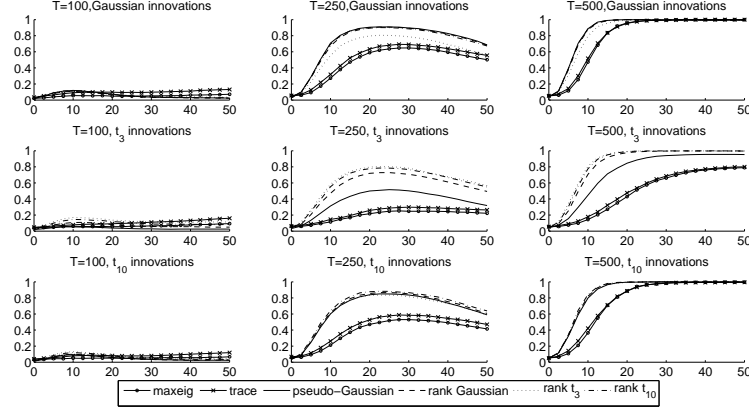


(a) for testing $H : r = 0$ versus $H' : r = 1$ under (5.2), for $h \in \{0, 2.5, 5, \dots, 50\}$, $p = 5$, $T \in \{100, 250, 500\}$, $\Sigma = \Sigma_{5,c}$, and $f \in \{\phi, t_3, t_{10}\}$;

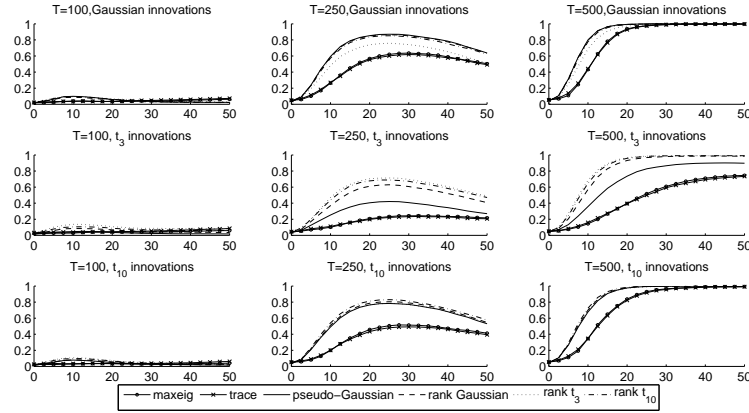


(b) for testing $H : r = 1$ versus $H' : r = 2$ under (5.2), for $h \in \{0, 2.5, 5, \dots, 50\}$, $p = 5$, $T \in \{100, 250, 500\}$, $\Sigma = \Sigma_{5,c}$, and $f \in \{\phi, t_3, t_{10}\}$.

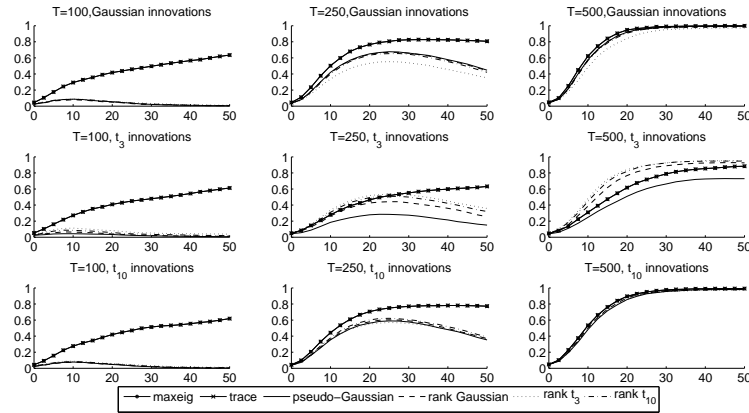
Figure 5.4: Simulated (2,500 replications) finite-sample powers of the *maxeig* test, the *trace* test, the pseudo-Gaussian test and the rank-based tests (4.9) with reference densities $q \in \{\phi, t_3, t_{10}\}$



(a) for testing $H : r = 2$ versus $H' : r = 3$ under (5.2), for $h \in \{0, 2.5, 5, \dots, 50\}$, $p = 5$, $T \in \{100, 250, 500\}$, $\Sigma = \Sigma_{5,c}$, and $f \in \{\phi, t_3, t_{10}\}$;



(b) for testing $H : r = 3$ versus $H' : r = 4$ under (5.2), for $h \in \{0, 2.5, 5, \dots, 50\}$, $p = 5$, $T \in \{100, 250, 500\}$, $\Sigma = \Sigma_{5,c}$, and $f \in \{\phi, t_3, t_{10}\}$;



(c) for testing $H : r = 4$ versus $H' : r = 5$ under (5.2), for $h \in \{0, 2.5, 5, \dots, 50\}$, $p = 5$, $T \in \{100, 250, 500\}$, $\Sigma = \Sigma_{5,c}$, and $f \in \{\phi, t_3, t_{10}\}$.

References

- BERNARD, A.B., AND S.N. DURLAF (1995). Convergence in international output. *J. of Applied Econometrics* 10, 97–108.
- BICKEL, P.J., C.A.J. KLAASSEN, Y. RITOV, AND J.A. WELLNER (1993). *Efficient and Adaptive Estimation for Semiparametric Models*. Johns Hopkins University Press, Baltimore, MD.
- BOSWIJK, H.P., M. JANSSEN, AND M.Ø. NIELSEN (2015). Improved likelihood ratio tests for cointegration rank in the VAR model. *J. Econometrics* 184, 97–110.
- BREITUNG (2001). Rank tests for nonlinear cointegration. *Journal of Business & Economic Statistics* 19, 331–340.
- BREITUNG, J., AND C. GOURIÉROUX (1997). Rank tests for unit roots. *J. Econometrics* 81, 7–27.
- DROST, F.C., C.A.J. KLAASSEN, AND B.J.M. WERKER (1997). Adaptive estimation in time series models. *Ann. Statist.* 25, 786–818.
- DROST, F.C., R. VAN DEN AKKER, AND B.J.M. WERKER (2009). The asymptotic structure of nearly unstable non negative integer-valued AR(1) models. *Bernoulli* 15, 297–324.
- ENGLE, R.F. AND C.W.J. GRANGER (1981). Co-integration and error correction: representation, estimation, and testing. *Econometrica* 55, 251–276.
- GRANGER, C.W.J. (1981). Some properties of time series data and their use in econometric model specification. *J. Econometrics* 16, 121–130.
- GUSHCHIN, A.A. (1996). Asymptotic optimality of parameter estimators under the LAQ condition. *Theory Probab. Appl.* 40, 261–272.
- HÁJEK, J. AND Z. ŠIDÁK (1967). *Theory of Rank Tests*. Academic Press, New York.
- HALL, P. AND C.C. HEYDE (1980). *Martingale Limit Theory and its Applications*. Academic Press, New York.
- HALLIN, M. AND CH. MEHTA (2015). R-estimation for asymmetric Independent Component Analysis. *Journal of the American Statistical Association*, to appear.
- HALLIN, M., H. OJA, AND D. PAINDAVEINE (2006). Semiparametrically efficient rank-based inference for shape. II. Optimal R-estimation of shape. *Ann. Statist.* 34, 2757–2789.
- HALLIN, M. AND D. PAINDAVEINE (2002a). Optimal tests for multivariate location based on interdirections and pseudo-Mahalanobis ranks. *Ann. Statist.* 30, 1103–1133.
- HALLIN, M. AND D. PAINDAVEINE (2002b). Optimal procedures based on interdirections and pseudo-Mahalanobis ranks for testing multivariate elliptic white noise against ARMA dependence. *Bernoulli* 8, 787–815.
- HALLIN, M. AND D. PAINDAVEINE (2005b). Affine-invariant aligned rank tests for the multivariate general linear model with ARMA errors. *J. Multivariate Anal.* 93, 122–163.

- HALLIN, M. AND D. PAINDAVEINE (2006). Asymptotic linearity of serial and nonserial multivariate signed rank statistics. *J. Statist. Plann. Inference* 136, 1–32.
- HALLIN, M., D. PAINDAVEINE, AND T. VERDEBOUT (2010). Optimal rank-based testing for principal components. *Ann. Statist.* 38, 3245–3299.
- HALLIN, M., D. PAINDAVEINE, AND T. VERDEBOUT (2013). Optimal rank-based tests for common principal components. *Bernoulli* 19, 2524–2556.
- HALLIN, M., R. VAN DEN AKKER, AND B.J.M. WERKER (2011). A class of simple semiparametrically efficient rank-based unit root tests. *J. Econometrics* 163, 200–214.
- HALLIN, M., R. VAN DEN AKKER, AND B.J.M. WERKER (2015). On quadratic expansions of log-likelihoods and a general asymptotic linearity result. In: *Mathematical Statistics and Limit Theorems : Festschrift in honour of Paul Deheuvels*, M. Hallin, D. M. Mason, D. Pfeifer, and J. Steinebach (eds.). Springer. Available at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2320807.
- HALLIN, M. AND B.J.M. WERKER (2003). Semiparametric efficiency, distribution-freeness, and invariance. *Bernoulli* 9, 137–165.
- HANSEN, B.E. (1992). Convergence to stochastic integrals for dependent heterogeneous processes. *Econometric Theory* 8, 489–500.
- HANSEN, P.R. (2005). Granger’s representation theorem: a closed-form expression for I(1) processes. *Econom. J.* 8, 23–28.
- HODGSON, D.J. (1998a). Adaptive estimation of cointegrating regressions with ARMA errors. *J. Econometrics* 85, 231–267.
- HODGSON, D.J. (1998b). Adaptive estimation of error correction models. *Econometric Theory* 14, 44–69.
- HODGSON, D.J., O. LINTON, AND K. VORKINK (2002). Testing the capital asset pricing model efficiently under elliptical symmetry: A semiparametric approach. *Journal of Applied Econometrics* 17, 617–639.
- HODGSON, D.J., AND K. VORKINK (2003). Efficient estimation of conditional asset pricing models. *Journal of Business and Economic Statistics* 21, 269–283.
- HUBRICH, K., H. LÜTKEPOHL, AND P. SAIKKONEN (2001). A review of systems cointegration tests. *Econometric Rev.* 20, 247–318.
- JACOD, J. AND A.N. SHIRYAEV (2002). *Limit Theorems for Stochastic Processes*. Berlin: Springer.
- JANSSON, M. (2008). Semiparametric power envelopes for tests of the unit root hypothesis. *Econometrica* 76, 1103–1142.
- JEGANATHAN, P. (1995). Some aspects of asymptotic theory with applications to time series models. *Econometric Theory* 11, 818–887.

- JEGANATHAN, P. (1997). On asymptotic inference in linear cointegrated time series systems. *Econometric Theory* 13, 692–745.
- JOHANSEN, S. (1988). Statistical analysis of cointegration vectors. *J. Econom. Dynam. Control* 12, 231–254.
- JOHANSEN, S. (1991). Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models. *Econometrica* 59, 1551–1580.
- JOHANSEN, S. (1995). *Likelihood-Based Inference in Cointegrated Vector Autoregressive Models*. New York: Oxford University Press.
- JOHANSEN, S. AND K. JUSELIUS (1990). Maximum likelihood estimation and inference on cointegration – with applications to the demand for money. *Oxford Bulletin of Economics and Statistics* 52, 169–210.
- KARATZAS, I. AND S.E. SHREVE (1991). *Brownian Motion and Stochastic Calculus*. Springer-Verlag, second edition.
- KREISS, J-P. (1987). On adaptative estimation in stationary ARMA processes. *Ann. Statist.* 15, 112–133.
- LE CAM, L. (1986). *Asymptotic Methods in Statistical Decision Theory*. New York: Springer-Verlag.
- LE CAM, L. AND G.L. YANG (1990). *Asymptotics in Statistics - some basic concepts*. New York: Springer.
- LEHMANN, E.L. AND J.P. ROMANO (2005). *Testing Statistical Hypothesis*. New York: Springer.
- MACKINNON, J. G., A. A. HAUG, AND L. MICHELIS (1999). Numerical distribution functions of Likelihood Ratio tests for cointegration. *J. Appl. Econometrics* 14, 563–577.
- MAGNUS, J.R. AND H. NEUDECKER (1988). *Matrix Differential Calculus with Applications in Statistics and Econometrics*. New York: John Wiley & Sons.
- MÜLLER, U.K. (2012). Efficient tests under a weak convergence assumption. *Econometrica* 79, 395–435.
- NASSEH, A. AND J. STRAUSS (2000). Stock prices and domestic and international macroeconomic activity: a cointegration approach. *Quart. Review of Economics and Finance* 40, 229–245.
- NIELSEN, B. (2009). Test for cointegration rank in general vector autoregressions. *Working Paper*, Department of Economics, University of Oxford.
- PHILLIPS, P.C.B. (1991). Optimal inference in cointegrated systems. *Econometrica* 59, 283–306.
- PLOBERGER, W. AND P.C.B. PHILLIPS (2012). Optimal estimation under nonstandard conditions. *J. Econometrics* 169, 258–265.
- RANDLES, R.H. (2000). A simpler, affine-invariant, multivariate, distribution-free sign test. *J. Amer. Statist. Assoc.* 95, 1263–1268.

- REINSEL, G.C. AND S.K. AHN (1992). Vector AR models with unit roots and reduced rank structure: estimation, likelihood ratio test, and forecasting. *J. Time Series Anal.* 13, 353–375.
- SAIKKONEN, P. AND H. LÜTKEPOHL (2000). Testing for the cointegrating rank of a VAR process with an intercept. *Econometric Theory* 16, 373–406.
- STOCK, J.H. (1987). Asymptotic properties of least squares estimators of cointegrating vectors. *Econometrica* 55, 1035–1056.
- STRASSER, H. (1985). *Mathematical Theory of Statistics: Statistical Experiments and Asymptotic Decision Theory*. Berlin: Walter de Gruyter.
- SWIFT, R. (2011). The relationship between health and GDP in OECD countries in the very long run. *Health Economics* 20, 306–322.
- TODA, H.Y. (1994). Finite sample properties of likelihood ratio tests for cointegrating ranks when linear trends are present. *Review of Economics and Statistics* 76, 66–79.
- TYLER, D.E. (1987). A distribution-free M-estimator of multivariate scatter. *Ann. Statist.* 15, 234–251.
- VAN DER VAART, A.W. (1988). *Statistical Estimation in Large Parameter Spaces*. CWI tract 44. Amsterdam: CWI.
- VAN DER VAART, A.W. (2000). *Asymptotic Statistics*. Cambridge: Cambridge University Press.
- WONG, T.W., M.C. CHIU, AND H.Y. WONG (2014). Time-consistent mean-variance hedging of longevity risk: effect of cointegration. *Insurance Mathematics and Economics* 56, 56–67.

Supplemental Appendix to “Optimal pseudo-Gaussian and rank-based tests of the cointegration rank in semiparametric error-correction Models”

A. Local Asymptotic Quadraticity for error-correction models

Introduce the filtrations $\mathbb{F}^{(T)} := \left(\mathcal{F}_u^{(T)}, u \in [0, 1] \right)$, $T \in \mathbb{N}$, defined by

$$\mathcal{F}_u^{(T)} := \sigma(\varepsilon_t, t \in \mathbb{N} : t \leq [uT]), \quad u \in [0, 1].$$

The angle-bracket process $\langle A_i^{(T)}, A_j^{(T)} \rangle(u)$ and the straight-bracket process $[A_i^{(T)}, A_j^{(T)}](u)$ are now well-defined for all $\mathbb{F}^{(T)}$ -adapted locally square-integrable martingales and semimartingales $A_i^{(T)}$, respectively (see, e.g., Jacod and Shiryaev (2002)). If $A_i^{(T)}$, $i = 1, 2$, are square-integrable martingales of the form $A_i^{(T)}(u) = \sum_{t=1}^{[uT]} I_{Tt}^{(i)}$ with $I_{Tt}^{(i)}$ \mathcal{F}_t -measurable, we have

$$[A_1^{(T)}, A_2^{(T)}](u) = \sum_{t=1}^{[uT]} I_{Tt}^{(1)} I_{Tt}^{(2)'} \quad \text{and} \quad \langle A_1^{(T)}, A_2^{(T)} \rangle(u) = \sum_{t=1}^{[uT]} \mathbb{E} \left[I_{Tt}^{(1)} I_{Tt}^{(2)'} \mid \mathcal{F}_{t-1} \right].$$

Recall that, for a square-integrable martingale with continuous sample paths, the angle-brackets and straight-brackets coincide.

Consider the ECM for the process $\{X_t\}$ as defined in (2.1). We parametrize local perturbations in the global parameters in such a way that we can identify exactly which parameter directions (which local subexperiments) lead to LAN, LAMN, or LABF behavior. For notational convenience, we introduce the following notation, with $\Psi_{\Gamma, \Pi}$ as defined in Assumption 3,

$$C = C_{\Gamma, \Pi} := \beta_{\perp} \Psi_{\Gamma, \Pi}^{-1} \alpha'_{\perp}.$$

Building on the factorization $\Pi = \alpha \beta'$, where α and β are full-rank $p \times r$ matrices, we define local (in the vicinity of $\vartheta = (\mu, \Gamma, \Pi)$) alternatives $\vartheta^{(T)} = (\mu^{(T)}, \Gamma^{(T)}, \Pi^{(T)})$, where

$$\mu^{(T)} = \mu_m^{(T)} := \mu + T^{-1/2} m, \quad \Gamma^{(T)} = \Gamma_G^{(T)} := \Gamma + T^{-1/2} G, \tag{A.1}$$

and

$$\Pi^{(T)} = \Pi_{A, b, B, d, D}^{(T)} := \alpha_A^{(T)} \beta_{b, B}^{(T)'} + T^{-3/2} \alpha_{\perp} d(C_{\Gamma, \Pi} \mu)' + T^{-1} \alpha_{\perp} D \beta'_{\perp}, \tag{A.2}$$

with local parameters $m \in \mathbb{R}^p$, $G_1, \dots, G_{k-1} \in \mathbb{R}^{p \times p}$, $d \in \mathbb{R}^{p-r}$, $D \in \mathbb{R}^{(p-r) \times (p-r)}$, and $\alpha_A^{(T)}$ and $\beta_{b,B}^{(T)}$ of the form

$$\alpha^{(T)} = \alpha_A^{(T)} := \alpha + \frac{A}{\sqrt{T}}, \quad \beta^{(T)} = \beta_{b,B}^{(T)} := \beta + \frac{(C_{\Gamma, \Pi} \mu) b'}{T^{3/2}} + \frac{\beta_{\perp} B'}{T}, \quad (\text{A.3})$$

for $A \in \mathbb{R}^{p \times r}$, $b \in \mathbb{R}^r$, and $B \in \mathbb{R}^{r \times (p-r)}$.

For the case $D = 0$, we are back to (3.1)-(3.3). The local parametrization $\Pi_{A,b,B,d,D}^{(T)}$ allows us (Proposition A.2 below) to derive a LAQ property that covers the ECM model with specified cointegration rank (corresponding to the restriction $d = 0 = D$) as well as the ECM model with unspecified cointegration rank (corresponding to local alternatives with $d \neq 0$, in case $\alpha'_{\perp} \mu \neq 0$, or $D \neq 0$). Note that the parametrization is such that both $d \neq 0$ and $D \neq 0$ increase the rank of Π ; in case $d \neq 0$ and $D = 0$, however, the rank of Π increases by one unit exactly.

The *central sequence* $\tilde{\Delta}_{\vartheta}^{(T)} = (\Delta_{\mu}^{(T)'}, \Delta_{\Gamma}^{(T)'}, \Delta_{\alpha}^{(T)'}, \Delta_b^{(T)'}, \Delta_B^{(T)'}, \Delta_d^{(T)'}, \Delta_D^{(T)'})'$, with $\Delta_{\Gamma}^{(T)} = (\Delta_{\Gamma_1}^{(T)'}, \dots, \Delta_{\Gamma_{k-1}}^{(T)'})'$, and *finite-sample Fisher information* $\tilde{J}_{\vartheta}^{(T)}$ appearing in Proposition A.2 below are

$$\tilde{\Delta}_{\vartheta}^{(T)} := \sum_{t=1}^T \begin{pmatrix} Z_{Tt}^{(1)}(\vartheta) \otimes I_p \\ Z_{Tt}^{(2)}(\vartheta) \otimes \alpha' \\ Z_{Tt}^{(2)}(\vartheta) \otimes \alpha'_{\perp} \end{pmatrix} \Sigma^{-1/2} U_t(\vartheta) \phi_f(\|\epsilon_t(\vartheta)\|_{\Sigma}), \quad (\text{A.4})$$

and

$$\tilde{J}_{\vartheta}^{(T)} := \frac{\mathcal{I}_p(f)}{p} \sum_{t=1}^T \begin{pmatrix} Z_{Tt}^{(1)} Z_{Tt}^{(1)'} \otimes \Sigma^{-1} & Z_{Tt}^{(1)} Z_{Tt}^{(2)'} \otimes \Sigma^{-1} \alpha & Z_{Tt}^{(1)} Z_{Tt}^{(2)'} \otimes \Sigma^{-1} \alpha_{\perp} \\ & Z_{Tt}^{(2)} Z_{Tt}^{(2)'} \otimes \alpha' \Sigma^{-1} \alpha & Z_{Tt}^{(2)} Z_{Tt}^{(2)'} \otimes \alpha' \Sigma^{-1} \alpha_{\perp} \\ & & Z_{Tt}^{(2)} Z_{Tt}^{(2)'} \otimes \alpha'_{\perp} \Sigma^{-1} \alpha_{\perp} \end{pmatrix}, \quad (\text{A.5})$$

respectively, where $Z_{Tt}^{(1)} = Z_{Tt}^{(1)}(\vartheta)$ and $Z_{Tt}^{(2)} = Z_{Tt}^{(2)}(\vartheta)$ as defined in (3.8)-(3.9).

In order to obtain convenient expressions for the *central sequence* $\tilde{\Delta}_{\vartheta}^{(T)}$ and the *finite-sample Fisher information* $\tilde{J}_{\vartheta}^{(T)}$, let us introduce some partial sum processes: for all $u \in [0, 1]$ and $\vartheta \in \Theta$, define (with $\epsilon_t = \epsilon_t(\vartheta)$, $U_t = U_t(\vartheta, \Sigma)$ and $Y_t = Y_t(\vartheta)$ as in (2.3))

$$W_{\epsilon}^{(T)}(u) := \frac{1}{\sqrt{T}} \sum_{t=1}^{[uT]} \epsilon_t, \quad W_{\phi}^{(T)}(u) := \frac{1}{\sqrt{T}} \sum_{t=1}^{[uT]} \Sigma^{-1/2} U_t \phi_f(\|\epsilon_t\|_{\Sigma}), \quad (\text{A.6})$$

$$W_{\Delta X \otimes \phi, j}^{(T)}(u) := \frac{1}{\sqrt{T}} \sum_{t=1}^{[uT]} \Delta X_{t-j} \otimes \Sigma^{-1/2} U_t \phi_f(\|\epsilon_t\|_{\Sigma}), \quad j = 1, \dots, k-1, \text{ and} \quad (\text{A.7})$$

$$W_{Y \otimes \phi}^{(T)}(u) := \frac{1}{\sqrt{T}} \sum_{t=1}^{[uT]} Y_{t-1} \otimes \Sigma^{-1/2} U_t \phi_f(\|\epsilon_t\|_{\Sigma}). \quad (\text{A.8})$$

Letting $m = p + p + (k-1)p^2 + p^2$, summarize the partial sum processes (A.6)-(A.8) in the process

$\mathcal{W}^{(T)}$ taking values in $D_{\mathbb{R}^m}[0, 1]$, i.e.

$$\mathcal{W}^{(T)'} = (W_{\epsilon}^{(T)'}, W_{\phi}^{(T)'}, W_{\Delta X \otimes \phi}^{(T)'}, W_{Y \otimes \phi}^{(T)'}) \quad \text{with} \quad W_{\Delta X \otimes \phi}^{(T)'} = (W_{\Delta X \otimes \phi, 1}^{(T)'}, \dots, W_{\Delta X \otimes \phi, k-1}^{(T)'}).$$

Let $\mathcal{W}' = (W'_{\epsilon}, W'_{\phi}, W'_{\Delta X \otimes \phi}, W'_{Y \otimes \phi})$ denote an m -variate Brownian motion with variance per unit of time

$$\begin{aligned} \text{Var} \begin{pmatrix} W_{\epsilon}(1) \\ W_{\phi}(1) \end{pmatrix} &= \begin{pmatrix} \frac{1}{p} \int_0^{\infty} r^{p+1} f(r) dr \Sigma & I_p \\ I_p & \frac{1}{p} \mathcal{I}_p(f) \Sigma^{-1} \end{pmatrix}, \\ \text{Var} \begin{pmatrix} W_{\Delta X \otimes \phi}(1) \\ W_{Y \otimes \phi}(1) \end{pmatrix} &= \frac{1}{p} \mathcal{I}_p(f) E_{\vartheta; \Sigma, f}^* \begin{pmatrix} \Delta X_0 \\ \vdots \\ \Delta X_{2-k} \\ Y_0 \end{pmatrix} \begin{pmatrix} \Delta X_0 \\ \vdots \\ \Delta X_{2-k} \\ Y_0 \end{pmatrix}' \otimes \Sigma^{-1}, \end{aligned}$$

and

$$\text{Cov} \left(\begin{pmatrix} W_{\Delta X \otimes \phi} \\ W_{Y \otimes \phi} \end{pmatrix}, \begin{pmatrix} W_{\epsilon}(1) \\ W_{\phi}(1) \end{pmatrix} \right) = \begin{pmatrix} 1_{(k-1) \times 1} \otimes C_{\Gamma, \Pi \mu} \\ E_{\vartheta; \Sigma, f}^* Y_0 \end{pmatrix} \otimes \begin{pmatrix} I_p & \frac{1}{p} \mathcal{I}_p(f) \Sigma^{-1} \end{pmatrix}.$$

Here $P_{\vartheta; \Sigma, f}^*$ denotes the probability measure under which the VAR process $\{V_t\}$, where

$$V'_t = ((\beta' X_t)', \Delta X'_t, \dots, \Delta X'_{t-k+2})$$

follows from the Granger-Johansen representation (2.3), is stationary, which is possible under Assumptions 1-3. The lemma below shows that $\mathcal{W}^{(T)}$ weakly converges to \mathcal{W} . The proof, an application of a functional central limit theorem for arrays of martingale differences, is included for completeness in Appendix C.1.

Lemma A.1. *Let Assumptions 1-3 hold, $\vartheta \in \Theta$, and $f \in \mathcal{F}_2$. Then, under $P_{\vartheta; \Sigma, f}^{(T)}$,*

$$(i) \quad \mathcal{W}^{(T)} \Rightarrow \mathcal{W} \text{ in } D_{\mathbb{R}^m}[0, 1], \text{ and} \tag{A.9}$$

$$(ii) \quad \langle \mathcal{W}^{(T)}, \mathcal{W}^{(T)} \rangle(1) = [\mathcal{W}^{(T)}, \mathcal{W}^{(T)}](1) + o_P(1) = \text{Var}(\mathcal{W}(1)) + o_P(1). \tag{A.10}$$

□

Using the Granger-Johansen representation (2.3), the central sequence $\tilde{\Delta}_{\vartheta}^{(T)}$ can also be expressed in terms of the partial sum processes (A.6)-(A.8):

$$\Delta_{\mu}^{(T)} = W_{\phi}^{(T)}(1), \quad \Delta_{\Gamma_j}^{(T)} = W_{\Delta X \otimes \phi, j}^{(T)}(1), \quad 1 \leq j \leq k-1, \quad \Delta_{\alpha}^{(T)} = (\beta' \otimes I_p) W_{Y \otimes \phi}^{(T)}(1),$$

$$\begin{aligned}
\Delta_b^{(T)} &= |C_{\Gamma, \Pi \mu}|^2 \int_0^1 \text{id}_T(u-) d(\alpha' W_\phi^{(T)})(u) + T^{-1/2} \int_0^1 (C_{\Gamma, \Pi \mu})' (C_{\Gamma, \Pi} W_\epsilon^{(T)}(u-)) d(\alpha' W_\phi^{(T)})(u) \\
&\quad + T^{-1} ((C_{\Gamma, \Pi \mu})' \otimes \alpha') W_{Y \otimes \phi}^{(T)}(1) + T^{-1} (C_{\Gamma, \Pi \mu})' a_{\Gamma, \Pi, \mu} \alpha' W_\phi^{(T)}(1), \\
\Delta_B^{(T)} &= \int_0^1 \beta'_\perp C_{\Gamma, \Pi} W_\epsilon^{(T)}(u-) \otimes I_r d(\alpha' W_\phi^{(T)})(u) + T^{-1} (\beta'_\perp \otimes \alpha') W_{Y \otimes \phi}^{(T)}(1) + T^{-1} (\beta'_\perp a_{\Gamma, \Pi, \mu} \otimes \alpha' W_\phi^{(T)}(1)), \\
\Delta_d^{(T)} &= |C_{\Gamma, \Pi \mu}|^2 \int_0^1 \text{id}_T(u-) d(\alpha'_\perp W_\phi^{(T)})(u) + T^{-1/2} \int_0^1 (C_{\Gamma, \Pi \mu})' (C_{\Gamma, \Pi} W_\epsilon^{(T)}(u-)) d(\alpha'_\perp W_\phi^{(T)})(u) \\
&\quad + T^{-1} ((C_{\Gamma, \Pi \mu})' \otimes \alpha'_\perp) W_{Y \otimes \phi}^{(T)}(1) + T^{-1} (C_{\Gamma, \Pi \mu})' a_{\Gamma, \Pi, \mu} \alpha'_\perp W_\phi^{(T)}(1), \\
\Delta_D^{(T)} &= \int_0^1 \beta'_\perp C_{\Gamma, \Pi} W_\epsilon^{(T)}(u-) \otimes I_{p-r} d(\alpha'_\perp W_\phi^{(T)})(u) + T^{-1} (\beta'_\perp \otimes \alpha'_\perp) W_{Y \otimes \phi}^{(T)}(1) + T^{-1} (\beta'_\perp a_{\Gamma, \Pi, \mu} \otimes \alpha'_\perp W_\phi^{(T)}(1)),
\end{aligned}$$

where id_T denotes the cadlag function $\text{id}_T(u) := [uT]/T$, which converges as $T \rightarrow \infty$ to the identity function $\text{id}(u) := u$ on $[0, 1]$.

We are now able to provide a complete characterization of the possible limiting local experiments in the ECM model (2.1). The proof is provided in Appendix C.2 and is based on an application of Proposition 1 in Hallin et al. (2015), which generalizes LAN results based on differentiability in quadratic mean (DQM) for models with independent observations to the time-series context with possibly non-LAN limits.

Proposition A.2. *Let Assumptions 1-3 hold, $\vartheta \in \Theta$, and $f \in \mathcal{F}_2$. Consider a bounded sequence of perturbations*

$$h'_T := (m'_T, (\text{vec } G_T)')', (\text{vec } A_T)')', b'_T, (\text{vec } B_T)')', d'_T, (\text{vec } D_T)')',$$

with $\beta_\perp B'_T \perp (C_{\Gamma, \Pi \mu})$, and $\beta_\perp D'_T \perp (C_{\Gamma, \Pi \mu})$, which defines a local parameter sequence $\vartheta^{(T)}$, see (A.1)-(A.3). Then,

(i) under $P_{\vartheta; \Sigma, f}^{(T)}$, as $T \rightarrow \infty$,

$$\log \frac{dP_{\vartheta^{(T)}; \Sigma, f}^{(T)}}{dP_{\vartheta; \Sigma, f}^{(T)}} = h'_T \tilde{\Delta}_{\vartheta}^{(T)} - \frac{1}{2} h'_T \tilde{J}_{\vartheta}^{(T)} h_T + o_P(1); \quad (\text{A.11})$$

(ii) still under $P_{\vartheta; \Sigma, f}^{(T)}$, $(\tilde{\Delta}_{\vartheta}^{(T)}, \tilde{J}_{\vartheta}^{(T)})$ converges in distribution to $(\tilde{\Delta}, \tilde{J})$ satisfying

$$\mathbb{E} \exp \left(h' \tilde{\Delta} - h' \tilde{J} h / 2 \right) = 1,$$

where $\tilde{\Delta}' := (\Delta'_\mu, \Delta'_{\Gamma_1}, \dots, \Delta'_{\Gamma_{k-1}}, \Delta'_\alpha, \Delta'_b, \Delta'_B, \Delta'_d, \Delta'_D)$ with, for $j = 1, \dots, k-1$,

$$\begin{aligned}
\Delta_\mu &:= W_\phi(1), \quad \Delta_{\Gamma_j} := W_{\Delta X \otimes \phi, j}(1), \quad \Delta_\alpha := (\beta' \otimes I_p) W_{Y \otimes \phi}(1), \\
\Delta_b &:= |C_{\Gamma, \Pi \mu}|^2 \int_0^1 u d(\alpha' W_\phi)(u), \quad \Delta_B := \int_0^1 (\beta'_\perp C_{\Gamma, \Pi} W_\epsilon(u) \otimes I_r) d(\alpha' W_\phi)(u), \\
\Delta_d &:= |C_{\Gamma, \Pi \mu}|^2 \int_0^1 u d(\alpha'_\perp W_\phi)(u), \quad \text{and} \quad \Delta_D := \int_0^1 (\beta'_\perp C_{\Gamma, \Pi} W_\epsilon(u) \otimes I_{p-r}) d(\alpha'_\perp W_\phi)(u),
\end{aligned}$$

respectively;

(iii) \tilde{J} conformally decomposes into blocks as

$$\tilde{J} = \frac{1}{p} \mathcal{I}_p(f) \begin{pmatrix} j_{11} & j_{12} \otimes \Sigma^{-1} \alpha & j_{12} \otimes \Sigma^{-1} \alpha_{\perp} \\ & j_{22} \otimes \alpha' \Sigma^{-1} \alpha & j_{22} \otimes \alpha' \Sigma^{-1} \alpha_{\perp} \\ & & j_{22} \otimes \alpha'_{\perp} \Sigma^{-1} \alpha_{\perp} \end{pmatrix},$$

with

$$j_{11} = \begin{pmatrix} I_p & 0 & 0 \\ 0 & I_{(k-1)p^2} & 0 \\ 0 & 0 & \beta' \otimes I_p \end{pmatrix} \text{Var} \begin{pmatrix} W_{\phi}(1) \\ W_{\Delta X \otimes \phi}(1) \\ W_{Y \otimes \phi}(1) \end{pmatrix} \begin{pmatrix} I_p & 0 & 0 \\ 0 & I_{(k-1)p^2} & 0 \\ 0 & 0 & \beta' \otimes I_p \end{pmatrix}',$$

$$j_{22} = \begin{pmatrix} \frac{1}{3} |C_{\Gamma, \Pi} \mu|^4 & |C_{\Gamma, \Pi} \mu|^2 \left(\int_0^1 u W_{\epsilon}(u) du \right)' C'_{\Gamma, \Pi} \beta_{\perp} \\ \beta'_{\perp} C_{\Gamma, \Pi} \left(\int_0^1 W_{\epsilon}(u) W'_{\epsilon}(u) du \right) C'_{\Gamma, \Pi} \beta_{\perp} \end{pmatrix},$$

and

$$j_{12} = \begin{pmatrix} \frac{1}{2} |C_{\Gamma, \Pi} \mu|^2 & \left(\int_0^1 W_{\epsilon}(u) du \right)' C'_{\Gamma, \Pi} \beta_{\perp} \\ 1_{(k-1) \times 1} \otimes \frac{1}{2} |C_{\Gamma, \Pi} \mu|^2 C_{\Gamma, \Pi} \mu & 1_{(k-1) \times 1} \otimes C_{\Gamma, \Pi} \mu \otimes \left(\int_0^1 W_{\epsilon}(u) du \right)' C'_{\Gamma, \Pi} \beta_{\perp} \\ \frac{1}{2} |C_{\Gamma, \Pi} \mu|^2 \beta' E_{\vartheta; \Sigma, f}^* Y_0 & \beta' E_{\vartheta; \Sigma, f}^* Y_0 \left(\int_0^1 W'_{\epsilon}(u) C'_{\Gamma, \Pi} \beta_{\perp} du \right) \end{pmatrix}.$$

□

Remark A.1.

- (i) Note that only the blocks involving B or D are random. Also note that all blocks involving b or d vanish if and only if $\alpha'_{\perp} \mu = 0$, i.e. when there is no deterministic linear trend.
- (ii) Proposition A.2 and Le Cam's first Lemma (see, e.g., Lemma 6.2 in Van der Vaart (2000)) jointly imply that the sequences of probability measures $P_{\vartheta^{(T)}; \Sigma, f}^{(T)}$ and $P_{\vartheta; \Sigma, f}^{(T)}$, $T \in \mathbb{N}$, are contiguous. Consequently, in expressions like (A.11), we do not have to worry whether α_P 's or O_P 's are taken at the null or at local alternatives of the form $P_{\vartheta^{(T)}; \Sigma, f}^{(T)}$. This consequence of contiguity is used throughout without further mention.
- (iii) The Brownian motions $C_{\Gamma, \Pi} W_{\epsilon}$ and $\alpha' W_{\phi}$ are independent, since

$$\mathbb{E} \left[C_{\Gamma, \Pi} \varepsilon_t \left(\alpha' \Sigma^{-1/2} U_t \phi_f(\|\varepsilon_t\|_{\Sigma}) \right)' \right] = \mathbb{E} [C_{\Gamma, \Pi} I_p \alpha] = 0$$

in view of the fact that $\mathbb{E} \|\varepsilon_t\|_{\Sigma} \phi_f(\|\varepsilon_t\|_{\Sigma}) = p$ and $C_{\Gamma, \Pi} \alpha = 0$.

- (iv) The Fisher Information matrix in Proposition A.2, both in finite-sample form $J_{\vartheta}^{(T)}$ and in the limit J , is somewhat involved. However, while its structure is used to classify the various limiting experiments, the exact forms are not needed in the rank-based test statistics of Section 4. \square

Remark A.2. In case $\alpha'_{\perp}\mu = 0_{p-r}$ the subexperiment corresponding to $B = 0$, $D = 0$, $d = 0$, leads to the LAMN result earlier obtained by Hodgson (1998b).¹⁸ It is somewhat surprising that, in case $\alpha'_{\perp}\mu \neq 0_{p-r}$, the subexperiment corresponding to $B = 0$ and $D = 0$ is not of the LAMN type. This is due to the fact that, conditionally on \tilde{J} , $\tilde{\Delta}$ is no longer zero-mean Gaussian due to, for instance, the non-zero correlation between W_{ϵ} and W_{ϕ} . A detailed study of the consequences of this phenomenon is beyond the scope of the present paper. Note that, with respect to the local parameters b and B only, we do obtain LAMN, as $\beta_{\perp}\Psi_{\Gamma,\Pi}^{-1}\alpha'_{\perp}W_{\epsilon}$ (which is the only source of randomness in $J_{b,B}$) and $\alpha'W_{\phi}$ are mutually independent. \square

B. Construction of locally discrete estimators

This appendix describes a refined discretization algorithm for β yielding the desired discreteness property while preserving the $T^{3/2}$ -consistency property of an initial estimator $\hat{\beta}^{(T)'}\beta_{\perp}\Psi_{\Gamma,\Pi}^{-1}\alpha'_{\perp}\mu$ (as an estimator of $\beta'\beta_{\perp}\Psi_{\Gamma,\Pi}^{-1}\alpha'_{\perp}\mu = 0$).

To be precise, consider an estimator $\hat{\beta}^{(T)}$ satisfying the consistency properties as in the statement of Assumption 5. For example, the reduced-rank regression estimator discussed in Johansen (1995) would do, see his Lemma 13.2. In a first step, round each element of $\hat{\beta}^{(T)}$ to a $T^{-3/2}$ -grid, that is, to the closest point in $\{kT^{-3/2} : k \in \mathbb{Z}\}$. In order not to clutter notation, we still denote this rounded estimator by $\hat{\beta}^{(T)}$, and observe that this first step preserves the T - and $T^{3/2}$ -consistency properties of the original $\hat{\beta}^{(T)}$. It also ensures that $T^{3/2}(\hat{\beta}^{(T)} - \beta)'\beta_{\perp}\Psi_{\Gamma,\Pi}^{-1}\alpha'_{\perp}\mu$, as $T \rightarrow \infty$, only has a bounded number of possible values within balls of radius M . However, $T(\hat{\beta}^{(T)} - \beta)$ in general still will take an unbounded number (of the order of $T^{1/2}$, to be precise) of possible values over such balls. Therefore, we apply a second discretization step, of order T , to each of the columns of $\hat{\beta}^{(T)}$, in such a way that the $T^{3/2}$ -consistency of $\hat{\beta}^{(T)'}\beta_{\perp}\Psi_{\Gamma,\Pi}^{-1}\alpha'_{\perp}\mu$ is preserved. Choose $j = 1, \dots, r$ and consider column $\hat{\beta}_j^{(T)}$. We now essentially project $\hat{\beta}_j^{(T)}$ sequentially $(p - r - 1)$ times on the closest of a series of parallel hyperplanes generated by the $(p - r - 1)$ columns of β_{\perp} that are orthogonal to $\beta_{\perp}\Psi_{\Gamma,\Pi}^{-1}\alpha'_{\perp}\mu$ and at distance of order T^{-1} . This

¹⁸For this specific parameter setting Hodgson (1998b) is more general, since that paper did not restrict to *elliptically* symmetric innovations.

general idea is complicated by the fact that both β_\perp and $\beta_\perp \Psi_{\Gamma, \Pi}^{-1} \alpha'_\perp \mu$ must be estimated, while the grid to be used is not allowed to be random.

For this second discretization of $\hat{\beta}_j^{(T)}$, note that, on the basis of Assumption 5, one can readily construct a $T^{1/2}$ -consistent estimator of rank $(p - r - 1)$ for $\beta_\perp - (\beta_\perp \Psi_{\Gamma, \Pi}^{-1} \alpha'_\perp \mu)(\beta_\perp \Psi_{\Gamma, \Pi}^{-1} \alpha'_\perp \mu)' \beta_\perp / |\beta_\perp \Psi_{\Gamma, \Pi}^{-1} \alpha'_\perp \mu|^2$. Choose vectors \hat{A}_l , $l = 1, \dots, p - r - 1$ that generate the same column space. Subsequently, discretize each of the components of \hat{A}_l on a $T^{-1/2}$ -grid. These discretized columns are, again for notational convenience, still denoted by \hat{A}_l . Note that $\hat{A}_l' \beta_\perp \Psi_{\Gamma, \Pi}^{-1} \alpha'_\perp \mu = O_P(T^{-1/2})$ under $P_{\vartheta; \Sigma, f}^{(T)}$. Now, perform in fact $(p - r - 1)$ discretization steps on $\hat{\beta}_j^{(T)}$ by sequentially projecting on the closest of the $(p - r - 1)$ hyperplanes $kT^{-1}\hat{A}_l + [\hat{A}_l]_\perp$, $l = 1, \dots, p - r - 1$, $k \in \mathbb{Z}$. The relevant insight is that such projections will not affect the $T^{3/2}$ -consistency of $\hat{\beta}_j^{(T)'} \beta_\perp \Psi_{\Gamma, \Pi}^{-1} \alpha'_\perp \mu$, as they affect $\hat{\beta}_j^{(T)}$ by a quantity $\eta \hat{A}_l$ for some $\eta \in (-1/T, 1/T)$ which, still under $P_{\vartheta; \Sigma, f}^{(T)}$, entails $\eta \hat{A}_l' \beta_\perp \Psi_{\Gamma, \Pi}^{-1} \alpha'_\perp \mu = O_P(T^{-3/2})$. Moreover, after all of these $(p - r - 1)$ projections have been carried out, the resulting $T(\hat{\beta}^{(T)} - \beta)$ (again, for notational simplicity, we keep the same notation for the discretized estimator as for the original one) only takes a bounded number of possible values over balls of radius M , due to the earlier $T^{1/2}$ -discretization of \hat{A}_l , $l = 1, \dots, p - r - 1$.

C. Proofs

C.1. Proof of Lemma A.1

Introduce $v_{t-1} := (\Delta X'_{t-1}, \dots, \Delta X'_{t-k+1}, Y'_{t-1})'$,

$$U_t^{(1)} := (\varepsilon'_t, (\Sigma^{-1/2} U_t^\varepsilon \phi_f(\|\varepsilon_t\|_\Sigma))')', \quad \text{and} \quad U_t^{(2)} := v_{t-1} \otimes \Sigma^{-1/2} U_t^\varepsilon \phi_f(\|\varepsilon_t\|_\Sigma).$$

Note that $\mathcal{W}^{(T)} = (\mathcal{W}_1^{(T)'}, \mathcal{W}_2^{(T)'})'$, with $\mathcal{W}_i^{(T)}(u) = T^{-1/2} \sum_{t=1}^{[uT]} U_t^{(i)}$, and partition $\mathcal{W} = (\mathcal{W}'_1, \mathcal{W}'_2)'$ accordingly. An application of Theorem VIII.3.33 in Jacod and Shiryaev (2002) shows that (A.9) and (A.10) hold if, for all $u \in [0, 1]$, $i, j \in \{1, 2\}$ and $\delta > 0$,

$$\langle \mathcal{W}_i^{(T)}, \mathcal{W}_j^{(T)} \rangle(u) = \frac{1}{T} \sum_{t=1}^{[uT]} \mathbb{E} \left[U_t^{(i)} U_t^{(j)'} \mid \mathcal{F}_{t-1} \right] \xrightarrow{P} u \text{Cov}(\mathcal{W}_i(1), \mathcal{W}_j(1)) \quad (\text{C.12})$$

and

$$\frac{1}{T} \sum_{t=1}^T \mathbb{E} \left[\left| U_t^{(i)} \right|^2 1 \left\{ \left| U_t^{(i)} \right| > \delta \sqrt{T} \right\} \mid \mathcal{F}_{t-1} \right] \xrightarrow{P} 0. \quad (\text{C.13})$$

We start with (C.12). Fix $u \in [0, 1]$. For $i = j = 1$, (C.12) is immediate from the weak law of large numbers (in view of the independence of U_t^ε and $\|\varepsilon_t\|_\Sigma$, and using the fact that $\mathbb{E}\|\varepsilon_t\|_\Sigma \phi_f(\|\varepsilon_t\|_\Sigma) = p$). For $i = j = 2$, we have

$$\frac{1}{T} \sum_{t=1}^{[uT]} \mathbb{E} \left[U_t^{(2)} U_t^{(2)'} \mid \mathcal{F}_{t-1} \right] = \frac{1}{p} \mathcal{I}_p(f) \left(\frac{1}{T} \sum_{t=1}^{[uT]} v_{t-1} v_{t-1}' \right) \otimes \Sigma^{-1}$$

and, for $i = 2$ and $j = 1$, we obtain

$$\frac{1}{T} \sum_{t=1}^{[uT]} \mathbb{E} \left[U_t^{(2)} U_t^{(1)'} \mid \mathcal{F}_{t-1} \right] = \frac{1}{T} \sum_{t=1}^{[uT]} v_{t-1} \otimes \text{Cov} \left(W_\phi(1), \begin{pmatrix} W_\varepsilon(1) \\ W_\phi(1) \end{pmatrix} \right).$$

Because $\{v_t\}$ is a stable vector autoregressive process of order 1 with finite second-order moments,

$$\frac{1}{T} \sum_{t=1}^{[uT]} v_{t-1} \xrightarrow{\text{P}} u \mathbb{E}_{\vartheta; \Sigma, f}^* \begin{pmatrix} \Delta X_0 \\ \vdots \\ \Delta X_{2-k} \\ Y_0 \end{pmatrix} \quad \text{and} \quad \frac{1}{T} \sum_{t=1}^{[uT]} v_{t-1} v_{t-1}' \xrightarrow{\text{P}} u \mathbb{E}_{\vartheta; \Sigma, f}^* \begin{pmatrix} \Delta X_0 \\ \vdots \\ \Delta X_{2-k} \\ Y_0 \end{pmatrix} \begin{pmatrix} \Delta X_0 \\ \vdots \\ \Delta X_{2-k} \\ Y_0 \end{pmatrix}'.$$

Note that $\Delta X_t = C_{\Gamma, \Pi} \mu + C_{\Gamma, \Pi} \varepsilon_t + \Delta Y_t$. We thus have $\mathbb{E}_{\vartheta; \Sigma, f}^* \Delta X_t = C_{\Gamma, \Pi} \mu$, which completes the verification of (C.12).

Next we verify (C.13). For $i = 1$, (C.13) is immediate (the L_1 norm of the left-hand-side converges to 0 by an application of the dominated convergence theorem). For $i = 2$, we first introduce

$$\phi(M) := \mathbb{E} \left| \Sigma^{-1/2} U_t^\varepsilon \phi_f(\|\varepsilon_t\|_\Sigma) \right|^2 \mathbb{1} \left\{ \left| \Sigma^{-1/2} U_t^\varepsilon \phi_f(\|\varepsilon_t\|_\Sigma) \right| \geq M \right\} \quad (\text{C.14})$$

and note that $\phi(M) \rightarrow 0$ as $M \rightarrow \infty$ (this essentially is the argument for $i = 1$). This yields

$$\frac{1}{T} \sum_{t=1}^T \mathbb{E} \left[\left| U_t^{(2)} \right|^2 \mathbb{1} \left\{ \left| U_t^{(2)} \right| > \delta \sqrt{T} \right\} \mid \mathcal{F}_{t-1} \right] \leq \phi \left(\frac{\delta \sqrt{T}}{\max_{t=1, \dots, T} |v_{t-1}|} \right) \frac{1}{T} \sum_{t=1}^T |v_{t-1}|^2 = o_{\text{P}}(1),$$

as we already noted that $T^{-1} \sum_{t=1}^T v_{t-1} v_{t-1}' = O_{\text{P}}(1)$, while $T^{-1/2} \max_{t=1, \dots, T} |v_{t-1}| = o_{\text{P}}(1)$ (this follows from a combination of (C.13) with Lemma 3 in Hallin et al. (2015)). This concludes the proof of Lemma A.1. \square

C.2. Proof of Proposition A.2

Throughout this section, expectations, O_{P} 's, and o_{P} 's are taken under $\text{P}_{\vartheta; \Sigma, f}$ —unless otherwise specified. To enhance readability, we split the proof into two parts. In Part A, we show that

$$(i) \ (\tilde{\Delta}_{\vartheta}^{(T)}, \tilde{J}_{\vartheta}^{(T)}) \Rightarrow (\tilde{\Delta}, \tilde{J}) \quad \text{and} \quad (ii) \ \mathbb{E} \exp(h' \tilde{\Delta} - h' \tilde{J} h / 2) = 1,$$

while Part B establishes the quadratic expansion (A.11) of log likelihood ratios.

Part A(i)

Consider the auxiliary processes (recall that id_T denotes the cadlag function $\text{id}_T(u) = [uT]/T$, which converges to the identity $\text{id}(u) = u$ on $[0, 1]$)

$$\tilde{\Delta}^{(T)}(u) := \begin{pmatrix} W_\phi^{(T)}(u) \\ W_{\Delta X \otimes \phi}^{(T)}(u) \\ (\beta' \otimes I_p) W_{Y \otimes \phi}^{(T)}(u) \\ |C_{\Gamma, \Pi} \mu|^2 \int_0^u \text{id}_T(s-) d(\alpha' W_\phi^{(T)})(s) \\ \int_0^u (\beta'_\perp C_{\Gamma, \Pi} W_\epsilon(s-) \otimes I_r) d(\alpha' W_\phi^{(T)})(s) \\ |C_{\Gamma, \Pi} \mu|^2 \int_0^u \text{id}_T(s-) d(\alpha'_\perp W_\phi^{(T)})(s) \\ \int_0^u (\beta'_\perp C_{\Gamma, \Pi} W_\epsilon(s-) \otimes I_{p-r}) d(\alpha'_\perp W_\phi^{(T)})(s) \end{pmatrix}, \quad T \in \mathbb{N}, \quad u \in [0, 1].$$

A combination of Lemma A.1 with Theorem 2.1 in Hansen (1992) (the conditions of which are trivially met, as $\mathcal{W}^{(T)}$ is a martingale with respect to $\mathbb{F}^{(T)}$ and the increments of $W_\phi^{(T)}$ are i.i.d. with finite second-order moment) yields $\tilde{\Delta}^{(T)} \Rightarrow \tilde{\Delta}$ in $D_{\mathbb{R}^{m^*}}[0, 1]$, where m^* denotes the number of components of $\tilde{\Delta}^{(T)}$. Using this weak convergence result, the identity

$$[A, B](u) = A(u)B(u) - A(0)B(0) - \int_0^u A(s-) dB(s) - \int_0^u B(s-) dA(s),$$

and the continuous mapping theorem, in combination again with Theorem 2.1 in Hansen (1992) (the conditions of which are met, here, since $\tilde{\Delta}^{(T)}$ is a martingale with respect to $\mathbb{F}^{(T)}$, and owing to the fact that $\sum_{t=1}^T \mathbb{E}|\tilde{\Delta}^{(T)}(t/T) - \tilde{\Delta}^{(T)}((t-1)/T)|^2 = O(1)$) yields

$$\left(\tilde{\Delta}^{(T)}(1), \left[\tilde{\Delta}^{(T)}, \tilde{\Delta}^{(T)} \right](1) \right) \Rightarrow \left(\tilde{\Delta}, \left[\tilde{\Delta}, \tilde{\Delta} \right](1) \right) = (\tilde{\Delta}, \tilde{J}). \quad (\text{C.15})$$

The representation (Page 3 of Appendix) of the central sequence $\tilde{\Delta}_\vartheta^{(T)}$ in terms of (integrals with respect to) $\mathcal{W}^{(T)}$ and the convergence result (C.15), along with the continuous mapping theorem, yield

$$\tilde{\Delta}_\vartheta^{(T)} = \hat{\Delta}^{(T)}(1) + o_P(1). \quad (\text{C.16})$$

Introducing $\tilde{Z}_{Tt} := (\tilde{Z}_{Tt}^{(1)'}, \tilde{Z}_{Tt}^{(2)'})'$, with $\tilde{Z}_{Tt}^{(1)} = Z_{Tt}^{(1)}$ and

$$\tilde{Z}_{Tt}^{(2)} = \frac{1}{\sqrt{T}} \begin{pmatrix} |C_{\Gamma, \Pi} \mu|^2 \frac{t-1}{T} \\ \beta'_\perp C_{\Gamma, \Pi} W_\epsilon^{(T)}\left(\frac{t-1}{T}\right) \end{pmatrix},$$

note that

$$\langle \tilde{\Delta}, \tilde{\Delta} \rangle(1) = \frac{\mathcal{I}_p(f)}{p} \sum_{t=1}^T \begin{pmatrix} \tilde{Z}_{Tt}^{(1)} \tilde{Z}_{Tt}^{(1)'}(\vartheta) \otimes \Sigma^{-1} & \tilde{Z}_{Tt}^{(1)} \tilde{Z}_{Tt}^{(2)'}(\vartheta) \otimes \Sigma^{-1} \alpha & \tilde{Z}_{Tt}^{(1)} \tilde{Z}_{Tt}^{(2)'}(\vartheta) \otimes \Sigma^{-1} \alpha_{\perp} \\ \tilde{Z}_{Tt}^{(2)} \tilde{Z}_{Tt}^{(2)'}(\vartheta) \otimes \alpha' \Sigma^{-1} \alpha & \tilde{Z}_{Tt}^{(2)} \tilde{Z}_{Tt}^{(2)'}(\vartheta) \otimes \alpha' \Sigma^{-1} \alpha_{\perp} & \\ \tilde{Z}_{Tt}^{(2)} \tilde{Z}_{Tt}^{(2)'}(\vartheta) \otimes \alpha'_{\perp} \Sigma^{-1} \alpha_{\perp} & & \end{pmatrix}.$$

In view of (C.15)-(C.16) and (A.5) the proof of Part A(i) is complete if we show that

$$\langle \tilde{\Delta}, \tilde{\Delta} \rangle(1) - \left[\tilde{\Delta}, \tilde{\Delta} \right](1) = o_P(1) \quad (\text{C.17})$$

and

$$\sum_{t=1}^T \tilde{Z}_{Tt} \tilde{Z}_{Tt}' - \sum_{t=1}^T Z_{Tt} Z_{Tt}' = o_P(1). \quad (\text{C.18})$$

Indeed, (C.18), and (A.5) imply $\tilde{J}_{\vartheta}^{(T)} = \langle \tilde{\Delta}, \tilde{\Delta} \rangle(1) + o_P(1)$ and (C.16)-(C.17) yield

$$\left(\tilde{\Delta}_{\vartheta}^{(T)}, \tilde{J}_{\vartheta}^{(T)} \right) = \left(\tilde{\Delta}^{(T)}(1), \left[\tilde{\Delta}, \tilde{\Delta} \right](1) \right) + o_P(1).$$

First consider (C.17). An application of Theorem 2.23 in Hall and Heyde (1980) shows that (C.17) holds if, for all $\delta > 0$, $i = 1, 2$ and compatible matrices M ,

$$\sum_{t=1}^T \mathbb{E} \left[\left| \tilde{Z}_{Tt}^{(i)} \otimes M \Sigma^{-1/2} U_t^{\varepsilon} \phi_f(\|\varepsilon_t\|_{\Sigma}) \right|^2 1\{|\tilde{Z}_{Tt}^{(i)} \otimes M \Sigma^{-1/2} U_t^{\varepsilon} \phi_f(\|\varepsilon_t\|_{\Sigma})| > \delta\} \mid \mathcal{F}_{t-1} \right] = o_P(1). \quad (\text{C.19})$$

We only show the result for $i = 2$ since, for $i = 1$, we already obtained the result for $M = I_p$ in the proof of Lemma A.1 (indeed, $\tilde{Z}_{Tt}^{(1)} = Z_{Tt}^{(1)}$). We easily obtain, with ϕ as defined in (C.14) and $\|\cdot\|$ denoting the Frobenius matrix norm, that the left-hand side of (C.19) is bounded by

$$\|M\|^2 Q_T \phi \left(\frac{\delta \sqrt{T}}{\|M\| \sqrt{Q_T}} \right),$$

with $Q_T = |C_{\Gamma, \Pi} \mu|^4 + p \|\beta'_{\perp} C_{\Gamma, \Pi}\|^2 \max_{i=1, \dots, p} \|W_{\varepsilon, i}^{(T)}\|_{\infty}^2$. Since $\|W_{\varepsilon}^{(T), i}\|_{\infty} \Rightarrow \|W_{\varepsilon, i}\|_{\infty}$, (C.19) follows from Lemma A.1 and the continuous mapping theorem.

Turning to (C.18), it follows (with $\tilde{Z}_{Tt}^{(1)} = Z_{Tt}^{(1)}$) from the proof of Lemma A.1 that

$$\sum_{t=1}^T Z_{Tt}^{(1)} Z_{Tt}^{(1)'} = O_P(1);$$

in view of the Cauchy-Schwarz inequality, it is thus sufficient to show that

$$\sum_{t=1}^T \left(\tilde{Z}_{Tt}^{(2)} \tilde{Z}_{Tt}^{(2)'} - Z_{Tt}^{(2)} Z_{Tt}^{(2)'} \right) = o_P(1).$$

This, however, easily follows from the decomposition

$$Z_{Tt}^{(2)} = \begin{pmatrix} (C_{\Gamma, \Pi} \mu)' \left(\frac{t-1}{T^{3/2}} C_{\Gamma, \Pi} \mu + \frac{1}{T} C_{\Gamma, \Pi} W_{\epsilon}^{(T)} \left(\frac{t-1}{T} \right) + \frac{1}{T^{3/2}} (Y_{t-1} + a_{\mu, \Gamma, \Pi}) \right) \\ \beta'_{\perp} \left(\frac{1}{\sqrt{T}} C_{\Gamma, \Pi} W_{\epsilon}^{(T)} \left(\frac{t-1}{T} \right) + \frac{1}{T} (Y_{t-1} + a_{\mu, \Gamma, \Pi}) \right) \end{pmatrix},$$

implied by the Granger-Johansen representation (2.3), in combination with the stability of the process $\{Y_t\}$, Lemma A.1, and the continuous mapping theorem—which completes the proof of Part A(i).

In Part B of the proof, we will exploit the fact that

$$\sum_{t=1}^T |Z_{Tt}|^2 = O_P(1) \quad \text{and} \quad \max_{t=1, \dots, T} |Z_{Tt}| = o_P(1), \quad (\text{C.20})$$

which follow from a combination of (C.17)-(C.18) with (C.15) and a combination of (C.18)-(C.19) with Lemma 3 in Hallin et al. (2015), respectively. \square

Part A(ii)

All components of the limiting central sequence Δ can be expressed as stochastic integrals with respect to linear combinations of \mathcal{W} , of the form $\Delta_i = \int_0^1 \xi_i(u) d\tilde{\mathcal{W}}_i$ (where $\tilde{\mathcal{W}}_i = a'_i \mathcal{W}$), with integrands ξ_i satisfying

$$\max_i \|\xi_i\|_{\infty} \leq C(1 + \max_i \|\mathcal{W}_i\|_{\infty}),$$

for some constant C (depending on ϑ). An application of Corollary 3.5.16 in Karatzas and Shreve (1991) yields $E \exp(h' \Delta - h' J h / 2) = 1$. \square

Part B

We use Proposition 1 in Hallin et al. (2015) to establish the validity of expansion (A.11). To this end, set $\tilde{P}_T := P_{\vartheta^{(T)}; \Sigma, f}^{(T)}$, $P_T := P_{\vartheta; \Sigma, f}^{(T)}$, and $\mathcal{F}_{Tt} := \mathcal{F}_t$. For notational convenience, we also introduce, for $T \in \mathbb{N}$ and $t = 1, \dots, T$,

$$S_{Tt} := \begin{pmatrix} Z_{Tt}^{(1)} \otimes \Sigma^{-1/2} U_t^{\epsilon} \phi_f(\|\epsilon_t\|_{\Sigma}) \\ Z_{Tt}^{(2)} \otimes \alpha' \Sigma^{-1/2} U_t^{\epsilon} \phi_f(\|\epsilon_t\|_{\Sigma}) \\ Z_{Tt}^{(2)} \otimes \alpha'_{\perp} \Sigma^{-1/2} U_t^{\epsilon} \phi_f(\|\epsilon_t\|_{\Sigma}) \end{pmatrix};$$

note that $\tilde{\Delta}_{\vartheta}^{(T)} = \sum_{t=1}^T S_{Tt}$ and $\tilde{J}_{\vartheta}^{(T)} = \sum_{t=1}^T E[S_{Tt} S'_{Tt} | \mathcal{F}_{t-1}]$.

In the notation of Proposition 1 of Hallin et al. (2015), we have

$$LR_{Tt} = \frac{f(\|\epsilon_t - w_{Tt}\|_{\Sigma})}{f(\|\epsilon_t\|_{\Sigma})}, \quad (\text{C.21})$$

with

$$\begin{aligned}
w_{Tt} = & T^{-1/2}m_T + T^{-1/2}\sum_{j=1}^{k-1}G_{T,j}\Delta X_{t-j} + T^{-1/2}A_T\beta'_T X_{t-1} + T^{-3/2}\alpha b_T(C_{\Gamma,\Pi}\mu)'X_{t-1} + T^{-1}\alpha B_T\beta'_\perp X_{t-1} \\
& + T^{-3/2}\alpha_\perp d_T(C_{\Gamma,\Pi}\mu)'X_{t-1} + T^{-1}\alpha_\perp D_T\beta'_\perp X_{t-1} + T^{-2}A_T b_T(C_{\Gamma,\Pi}\mu)'X_{t-1} + T^{-3/2}A_T B_T\beta'_\perp X_{t-1}.
\end{aligned}$$

Exploiting the orthogonality restrictions $\beta_\perp B'_T \perp (C_{\Gamma,\Pi}\mu)$ and $\beta_\perp D'_T \perp (C_{\Gamma,\Pi}\mu)$, we can rewrite the fifth, seventh, and last term of w_{Tt} as

$$\begin{aligned}
T^{-1}\alpha B_T\beta'_\perp X_{t-1} &= T^{-1}\alpha B_T\beta'_\perp (X_{t-1} - (t-1)C_{\Gamma,\Pi}\mu), \\
T^{-1}\alpha_\perp D_T\beta'_\perp X_{t-1} &= T^{-1}\alpha_\perp D_T\beta'_\perp (X_{t-1} - (t-1)C_{\Gamma,\Pi}\mu), \quad \text{and} \\
T^{-3/2}A_T B_T\beta'_\perp X_{t-1} &= T^{-3/2}A_T B_T\beta'_\perp (X_{t-1} - (t-1)C_{\Gamma,\Pi}\mu).
\end{aligned}$$

Assumption 2 implies, see Hallin and Paindaveine (2002a, Section 1), that the mapping $e \mapsto f^{1/2}(|e|)$ is differentiable in quadratic mean: namely,

$$\frac{\sqrt{f}(\|e - w\|_\Sigma)}{\sqrt{f}(\|e\|_\Sigma)} = 1 + \frac{1}{2} \left[\phi_f(\|e\|_\Sigma) w' \frac{\Sigma^{-1}e}{\|e\|_\Sigma} + r(e, w) \right],$$

where

$$\psi(\delta) = \sup_{w: |w| \leq \delta} \frac{1}{|w|^2} \text{Er}^2(\varepsilon_1, w) \rightarrow 0 \text{ as } \delta \rightarrow 0. \quad (\text{C.22})$$

Using the identity $\text{vec}(AXB) = (B' \otimes A) \text{vec}(X)$, we obtain (see (1) in Hallin et al. (2015))

$$\sqrt{LR_{Tt}} = 1 + \frac{1}{2} (h'_T S_{Tt} + R_{Tt}), \quad (\text{C.23})$$

with

$$\begin{aligned}
R_{Tt} = & r(\varepsilon_t, w_{Tt}) + (T^{-2}A_T b_T(C_{\Gamma,\Pi}\mu)'X_{t-1})' \phi_f(\|\varepsilon_t\|_\Sigma) \Sigma^{-1/2} U_t \\
& + (T^{-3/2}A_T B_T\beta'_\perp (X_{t-1} - (t-1)C_{\Gamma,\Pi}\mu))' \phi_f(\|\varepsilon_t\|_\Sigma) \Sigma^{-1/2} U_t.
\end{aligned} \quad (\text{C.24})$$

In order to conclude that expansion (A.11) holds, it is thus sufficient to check that Conditions (a)-(d) in Proposition 1 of Hallin et al. (2015) are satisfied.

Condition (a). That is h_T is bounded is an immediate consequence of the fact that, by assumption, h_T converges.

Condition (b) requires $\{S_{Tt}\}$ to be a P_T -square integrable martingale difference (with respect to \mathcal{F}_t) satisfying the conditional Lindeberg condition and with tight squared conditional moments (Displays (2)

and (3) in Hallin et al. (2015)). The square-integrability of S_{Tt} readily follows from the assumption that $f \in \mathcal{F}_2$. The centering condition (2) in Hallin et al. (2015) follows from the independence of U_t and $\phi_f(\|\varepsilon_t\|_\Sigma)$ and the fact that, under $P_{\vartheta;\Sigma,f}$, $E\phi_f(\|\varepsilon_t\|_\Sigma) = 0$. Turning to the conditional Lindeberg condition (3), we have, for $\delta > 0$ and with ϕ as defined in (C.14),

$$\sum_{t=1}^T E \left[(h'_T S_{Tt})^2 1_{\{|h'_T S_{Tt}| > \delta\}} | \mathcal{F}_{t-1} \right] \leq C |h_T|^2 \phi \left(\frac{\delta}{|h_T|_\infty C \max_{t=1,\dots,T} |Z_{Tt}|} \right) \sum_{t=1}^T |Z_{Tt}|^2,$$

with $C = 1 + \|\alpha\| + \|\alpha_\perp\|$. The desired Lindeberg condition then is a consequence of (C.20). Finally, by Part A(i) of the proof, $J_T = J_\vartheta^{(T)} \Rightarrow J$, so that Condition (b) is satisfied.

Condition (c). That condition consists in two asymptotic negligibility properties ((4) and (5) in Hallin et al. (2015)) of the remainders R_{Tt} . First consider (5). It follows from (C.20) that

$$\sum_{t=1}^T E \left[\left((T^{-2} A_T b_T (C_{\Gamma, \Pi} \mu)' X_{t-1})' \phi_f(\|\varepsilon_t\|_\Sigma) \Sigma^{-1/2} U_t \right)^2 | \mathcal{F}_{t-1} \right] = o_P(1)$$

and

$$\sum_{t=1}^T E \left[\left((T^{-3/2} A_T B_T \beta'_\perp (X_{t-1} - (t-1) C_{\Gamma, \Pi} \mu)' \phi_f(\|\varepsilon_t\|_\Sigma) \Sigma^{-1/2} U_t \right)^2 | \mathcal{F}_{t-1} \right] = o_P(1).$$

Along with (C.24), this yields

$$\begin{aligned} \sum_{t=1}^T E [R_{Tt}^2 | \mathcal{F}_{t-1}] &\leq 4 \sum_{t=1}^T E [r^2(\varepsilon_t, w_{Tt}) | \mathcal{F}_{t-1}] + 4 o_P(1) \\ &\leq 4\psi \left(\max_{t=1,\dots,T} |w_{Tt}| \right) \sum_{t=1}^T |w_{Tt}|^2 + 4 o_P(1). \end{aligned}$$

Note that w_{Tt} can be written as a linear transformation, with bounded coefficients, of Z_{Tt} (that is, we have $w_{Tt} = A^{(T)} Z_{Tt}$ with $\sup_{T \in \mathbb{N}} \|A^{(T)}\| < \infty$). Therefore, (C.20) implies

$$\sum_{t=1}^T |w_{Tt}|^2 = O_P(1) \text{ and } \max_{t=1,\dots,T} |w_{Tt}| = o_P(1),$$

so that

$$\sum_{t=1}^T E [R_{Tt}^2 | \mathcal{F}_{t-1}] = o_P(1),$$

which is exactly what (4) in Hallin et al. (2015)) requires, holds. As we assumed that the radial density f is strictly positive, (5) follows.

Condition (d) is an immediate consequence of the fact that initial values are deterministic. \square

C.3. Proof of Proposition 3.2

The efficient score for d is obtained as the residual of the regression, in the covariance structure $J_{\mu, \Gamma, \alpha, b, d}$, of the score for d on that for the nuisances m , G , A , and b . Let us show that this residual is indeed Δ_d^* given in (3.10). First, observe (using Appendix A) that

$$\Delta_d - \Delta_d^* = \left| \beta_{\perp} \Psi_{\Gamma, \Pi}^{-1} \alpha'_{\perp} \mu \right|^2 \int_{u=0}^1 u d[\alpha_{\perp} - [(I_p - P_{\alpha}) \alpha_{\perp}]]' W_{\phi}(u) + \left| \beta_{\perp} \Psi_{\Gamma, \Pi}^{-1} \alpha'_{\perp} \mu \right|^2 \frac{1}{2} [(I_p - P_{\alpha}) \alpha_{\perp}]' W_{\phi}(1).$$

The components of the first term are in the space spanned by the components of the score for b , those of the second term in the space spanned by the components of the score for m . Since $\int_{u=0}^1 (u - 1/2) du = 0$, (3.10) moreover is orthogonal to $W_{\Delta X \otimes \phi}(1)$ and $W_{Y \otimes \phi}(1)$, thus to the scores induced by m , G , and A . Finally, (3.10) is also orthogonal to the scores induced by b , as $[(I_p - P_{\alpha}) \alpha_{\perp}]' W_{\phi}(u)$ and $\alpha' W_{\phi}(u)$ are independent (their covariance vanishes due to the fact that $[(I_p - P_{\alpha}) \alpha_{\perp}]' \Sigma^{-1} \alpha = 0$). In case $r_0 = 0$, the result follows along the same lines for the scores induced by m and G , while those induced by A and b need not be considered. This establishes Part (i) of the lemma. Part (ii) is an immediate consequence of the weak convergence results in Appendix A and the corresponding proofs. \square

C.4. Proofs Section 4

C.4.1. Proof of Proposition 4.1

The proof follows along the same lines as the (trickier) proofs of Propositions 4.2-4.3, and therefore is omitted. \square

C.5. Proof of Proposition 4.2

We organize the proof of each part in separate sections.

C.5.1. Proof of Part (i)

Part (i) of the proposition follows as a particular case of the nonserial asymptotic representation result in Proposition 2.1(i) of Hallin and Paindaveine (2006), with the vector-valued (deterministic) weights $\Sigma^{-1/2} x'_{t-i} K$ replaced with the scalar ones $\left(\frac{t}{T+1} - \frac{1}{2} \right)$. \square

C.5.2. Proof of Part (ii)

Part (ii) is a direct consequence of the central limit theorem for weighted sums of independent summands with finite variance and weights satisfying the traditional Noether condition (see Hájek and Sidák (1967), p.153), which directly applies to $\underline{S}_g^{(T)}(\vartheta; \Sigma)$ defined in (4.6). \square

C.5.3. Proof of Part (iii)

The asymptotic normality of $\underline{S}_g^{(T)}(\vartheta; \hat{\Sigma}^{(T)})$ in Part (iii) results from a classical application of Le Cam's third Lemma (see, for instance, p.90 of Van der Vaart (2000)). Due to contiguity, the asymptotic representation (4.6) also holds under $P_{\vartheta^{(T)}; \Sigma, f}^{(T)}$. Hence, the asymptotic mean of $\underline{S}_g^{(T)}(\vartheta; \hat{\Sigma}^{(T)})$, under local alternatives of the form $P_{\vartheta^{(T)}; \Sigma, f}^{(T)}$, is the vector of asymptotic covariances, under $P_{\vartheta; \Sigma, f}^{(T)}$, of the asymptotically joint normal distribution of $\underline{S}_g^{(T)}(\vartheta; \Sigma)$ in (4.6) and the log-likelihood (3.6). Clearly, those covariances involve sums of expectations of products of quantities of the type $U_t^\varepsilon U_{t'}^{\varepsilon'}$ and the corresponding summands in the linear part $h' \Delta_\vartheta^{(T)}$ of the approximation of local log-likelihoods; they break into $k + 3$ parts, associated with the corresponding subvectors of $\Delta_\vartheta^{(T)}$, which we successively examine.

(a) For the $\Delta_\mu^{(T)}$ part of $\Delta_\vartheta^{(T)}$, the covariance is

$$T^{-1} \sum_{t=1}^T \left(\frac{t}{T+1} - \frac{1}{2} \right) E \left[\phi_g(\tilde{G}_p^{-1}(\tilde{F}_p(\|\varepsilon_t\|_\Sigma))) \phi_f(\|\varepsilon_t\|_\Sigma) U_t^\varepsilon U_t^{\varepsilon'} \right] m = T^{-1} \sum_{t=1}^T \left(\frac{t}{T+1} - \frac{1}{2} \right) \frac{\mathcal{I}(f, g)}{p} m = 0.$$

(b) For the $\Delta_{\Gamma_i}^{(T)}$, $i = 1, \dots, k-1$ subvectors, we obtain the covariances

$$T^{-1} \sum_{t=1}^T \left(\frac{t}{T+1} - \frac{1}{2} \right) E \left[\phi_g(\tilde{G}_p^{-1}(\tilde{F}_p(\|\varepsilon_t\|_\Sigma))) \phi_f(\|\varepsilon_t\|_\Sigma) U_t^\varepsilon U_t^{\varepsilon'} \Sigma^{-1/2} (\Delta X'_{t-i} \otimes I_p) \right] \text{vec}(G_i)$$

where $\Delta X_{t-i} = C_{\Pi, \Gamma} \mu + C_{\Pi, \Gamma} \varepsilon_{t-i} + \Delta Y_{t-i}$ is independent of ε_t and U_t^ε , yielding

$$T^{-1} \sum_{t=1}^T \left(\frac{t}{T+1} - \frac{1}{2} \right) \frac{\mathcal{I}(f, g)}{p} \Sigma^{-1/2} \left((C_{\Pi, \Gamma} \mu + E[\Delta Y_{t-i}]) \otimes I_p \right) \text{vec}(G_i) = o(1)$$

since $\sum_{t=1}^T \left(\frac{t}{T+1} - \frac{1}{2} \right) = 0$ and $E[\Delta Y_{t-i}] = o(1)$ in view of the asymptotic stationarity of $\{Y_t\}$.

(c) Turning to Δ_α , the covariance term is

$$T^{-1} \sum_{t=1}^T \left(\frac{t}{T+1} - \frac{1}{2} \right) E \left[\phi_g(\tilde{G}_p^{-1}(\tilde{F}_p(\|\varepsilon_t\|_\Sigma))) \phi_f(\|\varepsilon_t\|_\Sigma) U_t^\varepsilon U_t^{\varepsilon'} \Sigma^{-1/2} ((X'_{t-1} \beta) \otimes I_p) \right] \text{vec}(A),$$

which, due to the fact that $X'_{t-1} \beta = Y'_{t-1} \beta$, where Y_{t-1} is independent of ε_t , takes the form

$$T^{-1} \sum_{t=1}^T \left(\frac{t}{T+1} - \frac{1}{2} \right) \frac{\mathcal{I}(f, g)}{p} \Sigma^{-1/2} \left((C_{\Pi, \Gamma} \mu + E[Y'_{t-1} \beta]) \otimes I_p \right) \text{vec}(A) = o(1)$$

since, again, $\sum_{t=1}^T \left(\frac{t}{T+1} - \frac{1}{2} \right) = 0$ and $\{Y_t\}$ is asymptotically stationary.

(d) The perturbation of β' has two parts, one with rate $T^{-3/2}$, the other one with rate T^{-1} . The first one yields a covariance term

$$T^{-2} \sum_{t=1}^T \left(\frac{t}{T+1} - \frac{1}{2} \right) \mathbb{E} \left[\phi_g(\tilde{G}_p^{-1}(\tilde{F}_p(\|\varepsilon_t\|_\Sigma))) \phi_f(\|\varepsilon_t\|_\Sigma) U_t^\varepsilon U_t^{\varepsilon'} \Sigma^{-1/2} (X'_{t-1} \otimes \alpha) \right] \text{vec}(b(C_{\Gamma, \Pi} \mu)')$$

which, taking into account the Granger representation (2.3) of X_{t-1} and the independence between X_{t-1} and ε_t (hence also U_t^ε), reduces to

$$T^{-2} \sum_{t=1}^T \left(\frac{t}{T+1} - \frac{1}{2} \right) \frac{\mathcal{I}(f, g)}{p} \Sigma^{-1/2} \left(((t-1)C_{\Pi, \Gamma} \mu)' + \mathbb{E}[Y'_{t-1}] + a'_{\Pi, \Gamma, \mu} \right) \otimes \alpha \text{vec}(b(C_{\Gamma, \Pi} \mu)') \\ = A_1 + A_2 + A_3, \quad \text{say.}$$

Now, recalling that $(C' \otimes A) \text{vec}(B) = \text{vec}(ABC)$,

$$A_1 = T^{-2} \sum_{t=1}^T \left(\frac{t}{T+1} - \frac{1}{2} \right) (t-1) \frac{\mathcal{I}(f, g)}{p} \Sigma^{-1/2} ((C_{\Pi, \Gamma} \mu)' \otimes \alpha) \text{vec}(b(C_{\Gamma, \Pi} \mu)') \\ = \frac{\mathcal{I}(f, g)}{12p} \Sigma^{-1/2} \text{vec}(\alpha b(C_{\Pi, \Gamma} \mu)' (C_{\Pi, \Gamma} \mu)) = \frac{\mathcal{I}(f, g)}{12p} \|C_{\Gamma, \Pi} \mu\|^2 \Sigma^{-1/2} \alpha b, \quad (\text{C.25})$$

while, in view of $\{Y_t\}$'s asymptotic stationarity,

$$A_2 = T^{-2} \sum_{t=1}^T \left(\frac{t}{T+1} - \frac{1}{2} \right) \frac{\mathcal{I}(f, g)}{p} \Sigma^{-1/2} (\mathbb{E}[Y'_{t-1}]) \otimes \alpha \text{vec}(b(C_{\Gamma, \Pi} \mu)') = o(1)$$

as $T \rightarrow \infty$ and, since $\sum_{t=1}^T \left(\frac{t}{T+1} - \frac{1}{2} \right) = 0$,

$$A_3 = T^{-2} \sum_{t=1}^T \left(\frac{t}{T+1} - \frac{1}{2} \right) \frac{\mathcal{I}(f, g)}{p} \Sigma^{-1/2} (a'_{\Pi, \Gamma, \mu} \otimes \alpha) \text{vec}(b(C_{\Gamma, \Pi} \mu)') = 0.$$

The T^{-1} -perturbation of β' is of the form $T^{-1}B\beta'_\perp$, where $B\beta'_\perp C_{\Pi, \Gamma} \mu = 0$. Its contribution to the covariance decomposes, as the previous one, into a sum of three terms, A'_1 , A'_2 and A'_3 , say, with, for the same reasons as above, $A'_2 = o(1)$ and $A'_3 = 0$. As for A'_1 , it takes the form

$$A'_1 = T^{-3/2} \sum_{t=1}^T \left(\frac{t}{T+1} - \frac{1}{2} \right) (t-1) \frac{\mathcal{I}(f, g)}{p} \Sigma^{-1/2} ((C_{\Pi, \Gamma} \mu)' \otimes \alpha) \text{vec}(B\beta'_\perp) \\ = T^{-3/2} \sum_{t=1}^T \left(\frac{t}{T+1} - \frac{1}{2} \right) (t-1) \frac{\mathcal{I}(f, g)}{p} \Sigma^{-1/2} \text{vec}(\alpha B\beta'_\perp C_{\Pi, \Gamma} \mu)$$

where $T^{-3/2} \sum_{t=1}^T \left(\frac{t}{T+1} - \frac{1}{2} \right) (t-1)$ does not converge anymore; the orthogonality $B\beta'_\perp C_{\Pi, \Gamma} \mu = 0$, however, implies that $A'_1 = 0$, irrespective of T .

(e) The reasoning for the $\Delta_d^{(T)}$ -part of the central sequence is entirely similar, and yields a contribution

$$\frac{\mathcal{I}(f, g)}{12p} \|C_{\Gamma, \Pi\mu}\|^2 \Sigma^{-1/2} \alpha_\perp d; \quad (\text{C.26})$$

details are left to the reader.

The desired result follows from adding the contributions (C.25) and (C.26). \square

C.5.4. Proof of Part (iv)

The proof of the asymptotic linearity (4.8) in Part (iv) of Proposition 4.2 relies on Proposition 2 in Hallin et al. (2015). In view of the asymptotic equivalence (4.6) and contiguity, it is sufficient, in order to establish (4.8), to show that, under $P_{\vartheta; \Sigma, f}^{(T)}$,

$$\underline{S}_g^{(T)}(\vartheta^{(T)}; \Sigma) - \underline{S}_g^{(T)}(\vartheta; \Sigma) = -\frac{1}{p} \mathcal{I}_p(f, g) \frac{1}{\sqrt{T}} \sum_{t=1}^T \left(\frac{t}{T+1} - \frac{1}{2} \right) v_{Tt} + o_P(1). \quad (\text{C.27})$$

Moreover, in view of Assumption 6, it is sufficient to prove (C.27) for monotone increasing ϕ_g .

Introduce, for $m \in \mathbb{N}$, the truncated versions J_m of $\phi_g \circ \tilde{G}_p^{-1}$

$$J_m(u) := \begin{cases} 0 & \text{if } 0 \leq u \leq \frac{1}{m}; \\ \phi_g \circ \tilde{G}_p^{-1} \left(\frac{2}{m} \right) m \left(u - \frac{1}{m} \right) & \text{if } \frac{1}{m} < u \leq \frac{2}{m}; \\ \phi_g \circ \tilde{G}_p^{-1}(u) & \text{if } \frac{2}{m} < u \leq 1 - \frac{2}{m}; \\ \phi_g \circ \tilde{G}_p^{-1} \left(1 - \frac{2}{m} \right) m \left(1 - \frac{1}{m} - u \right) & \text{if } 1 - \frac{2}{m} < u \leq 1 - \frac{1}{m}; \\ 0 & \text{if } 1 - \frac{1}{m} \leq u \leq 1. \end{cases}$$

Note that $J_m : [0, 1] \rightarrow \mathbb{R}$ is uniformly continuous and thus bounded. Moreover, we have

$$\int_0^1 (\phi_g \circ \tilde{G}_p^{-1}(u))^2 du < \infty, \quad \lim_{m \rightarrow \infty} J_m(u) = \phi_g \circ \tilde{G}_p^{-1}(u), \quad u \in (0, 1), \quad (\text{C.28})$$

and, using the fact that $\phi_g \circ \tilde{G}_p^{-1}$ is non-decreasing and continuous, for some $m_0 \in \mathbb{N}$,

$$|J_m(u)| \leq \left| \phi_g \circ \tilde{G}_p^{-1}(u) \right|, \quad u \in (0, 1), \quad m \geq m_0. \quad (\text{C.29})$$

Next, consider, for $\theta \in \Theta$, the truncated-score statistics

$$\underline{S}_{g;m}^{(T)}(\theta; \Sigma) := \frac{1}{\sqrt{T}} \sum_{t=1}^T \left(\frac{t+1}{T} - \frac{1}{2} \right) U_t(\theta, \Sigma) J_m(\tilde{F}_p(\|\epsilon_t(\theta)\|_\Sigma)).$$

The following lemma relates these truncated-score statistics to their non-truncated counterparts.

Lemma C.1. For all $\epsilon > 0$, $\lim_{m \rightarrow \infty} \sup_{T \in \mathbb{N}} \sup_{\vartheta \in \Theta} \mathbb{P}_{\vartheta; \Sigma, f} \left(\left| \underline{S}_g^{(T)}(\vartheta; \Sigma) - \underline{S}_{g; m}^{(T)}(\vartheta; \Sigma) \right| > \epsilon \right) = 0.$ \square

The proof of this lemma in turn involves another lemma.

Lemma C.2. Under $\mathbb{P}_{\vartheta; \Sigma, f}$,

$$\max_{1 \leq t \leq T} \left| \|\epsilon_t(\vartheta^{(T)})\|_{\Sigma} - \|\epsilon_t(\vartheta)\|_{\Sigma} \right| = o_{\mathbb{P}}(1), \quad (\text{C.30})$$

and, for all $\eta > 0$,

$$\max_{1 \leq t \leq T} |U_t(\vartheta^{(T)}, \Sigma) - U_t(\vartheta, \Sigma)| 1\{\|\epsilon_t\|_{\Sigma} > \eta\} = o_{\mathbb{P}}(1). \quad (\text{C.31})$$

\square

Proof. Under $\mathbb{P}_{\vartheta; \Sigma, f}$, we have (see the proof of Lemma 4.1 in Hallin and Paindaveine (2006, pp.29-30) for details),

$$\left| \|\epsilon_t(\vartheta^{(T)})\|_{\Sigma} - \|\epsilon_t(\vartheta)\|_{\Sigma} \right| \leq \|\Sigma^{-1/2}\| \|\epsilon_t(\vartheta) - \epsilon_t(\vartheta^{(T)})\|$$

and

$$|U_t(\vartheta^{(T)}, \Sigma) - U_t(\vartheta, \Sigma)| 1\{\|\epsilon_t\|_{\Sigma} > \eta\} \leq \frac{2}{\eta} \|\Sigma^{-1/2}\| \|\epsilon_t(\vartheta) - \epsilon_t(\vartheta^{(T)})\|.$$

Hence, in order to prove that (C.30) and (C.31) hold, it is sufficient to show that

$$\max_{t=1, \dots, T} |\epsilon_t(\vartheta) - \epsilon_t(\vartheta^{(T)})| = o_{\mathbb{P}}(1). \quad (\text{C.32})$$

We have

$$\begin{aligned} \epsilon_t(\vartheta^{(T)}) - \epsilon_t &= -T^{-1/2} m_T - \frac{1}{\sqrt{T}} \sum_{j=1}^{k-1} G_j^{(T)} \Delta X_{t-j} - \frac{1}{\sqrt{T}} A_T \beta' Y_{t-1} \\ &\quad - T^{-3/2} (t-1) |C_{\Gamma, \Pi} \mu|^2 \alpha^{(T)} b_T - T^{-1} \alpha^{(T)} b_T (C_{\Gamma, \Pi} \mu)' C_{\Gamma, \Pi} W_{\epsilon}^{(T)} \left(\frac{t-1}{T} \right) \\ &\quad - T^{-3/2} \alpha^{(T)} b_T (C_{\Gamma, \Pi} \mu)' (Y_{t-1} + a_{\mu, \Gamma, \Pi}) - T^{-1} \alpha^{(T)} B_T \beta'_{\perp} C_{\Gamma, \Pi} W_{\epsilon}^{(T)} \left(\frac{t-1}{T} \right) \\ &\quad - T^{-1} \alpha^{(T)} B_T \beta'_{\perp} (Y_{t-1} + a_{\mu, \Gamma, \Pi}). \end{aligned}$$

As the process $\{V_t\}$, from the Granger-Johansen representation (2.3), is, under $\mathbb{P}_{\vartheta; \Sigma, f}$, stable with finite second moments we have, for all $\delta > 0$,

$$\lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T \mathbb{E}_{\vartheta; \Sigma, f} |V_t|^2 1\{|V_t| > \delta \sqrt{T}\} = 0,$$

which implies, by an application of Lemma 3 in Hallin et al. (2015), that $T^{-1/2} \max_{t=1, \dots, T} |V_t| = o_P(1)$. It follows that $T^{-1/2} \max_{t=1, \dots, T} |\Delta X_{t-j}|$ ($j = 1, \dots, k-1$) and $T^{-1/2} \max_{t=1, \dots, T} |Y_t|$ are $o_P(1)$. Now, an application of Lemma A.1 and the continuous mapping theorem yield $\max_{i=1, \dots, p} \|W_{\epsilon, i}^{(T)}\|_\infty = O_P(1)$. This establishes (C.32) and hence concludes the proof of Lemma C.2. \square

Proof of Lemma C.1. Let

$$C_{t;m}^{(T)} := U_t^\varepsilon (\phi_g \circ \tilde{G}_p^{-1} \circ \tilde{F}_p(\|\varepsilon_t\|_\Sigma) - J_m(\tilde{F}_p(\|\varepsilon_t\|_\Sigma))).$$

Then, we have

$$\begin{aligned} \text{Var}_{\vartheta; \Sigma, f}(\underline{S}_g^{(T)}(\vartheta; \Sigma) - \underline{S}_{g;m}^{(T)}(\vartheta; \Sigma)) &= \frac{1}{T} \sum_{t=1}^T \left(\frac{t}{T+1} - \frac{1}{2} \right)^2 \mathbb{E}_{\vartheta; \Sigma, f}[C_{t;m}^{(T)} C_{t;m}^{(T)'}] \\ &= \frac{1}{p} I_p \mathbb{E}(\phi_g \circ \tilde{G}_p^{-1}(V) - J_m(V))^2 \frac{1}{T} \sum_{t=1}^T \left(\frac{t}{T+1} - \frac{1}{2} \right)^2, \end{aligned}$$

where V is uniform over $[0, 1]$. Dominated convergence and (C.28)-(C.29) yield that $\mathbb{E}[(\phi_g \circ \tilde{G}_p^{-1}(V) - J_m(V))^2]$, which does not depend on T and ϑ , converges to zero as $m \rightarrow \infty$. Since $\underline{S}_g^{(T)}(\vartheta; \Sigma)$ and $\underline{S}_{g;m}^{(T)}(\vartheta; \Sigma)$ are both centered under $\mathbb{P}_{\vartheta; \Sigma, f}^{(T)}$, an application of the Markov inequality completes the proof of Lemma C.1. \square

We now turn back to the proof of part (iv) of Proposition 4.2. Defining

$$\mathcal{I}_{p,m}(f, g) := \int_0^1 \phi_f(\tilde{F}_p^{-1}(u)) J_m(u) du,$$

note that dominated convergence, (C.28) and (C.29) entail

$$\lim_{m \rightarrow \infty} \mathcal{I}_{p,m}(f, g) = \mathcal{I}_p(f, g). \quad (\text{C.33})$$

As $T^{-1/2} \sum_{t=1}^T (t/(T+1) - 1/2) v_{Tt} = O_P(1)$, Lemma C.1, contiguity, and (C.33) imply that a sufficient condition for (C.27) to hold is that, for all $m \in \mathbb{N}$,

$$\underline{S}_{g;m}^{(T)}(\vartheta^{(T)}; \Sigma) - \underline{S}_{g;m}^{(T)}(\vartheta; \Sigma) = -\frac{1}{p} \mathcal{I}_{p,m}(f, g) \frac{1}{\sqrt{T}} \sum_{t=1}^T \left(\frac{t}{T+1} - \frac{1}{2} \right) v_{Tt} + o_P(1). \quad (\text{C.34})$$

Let us show that (C.34) holds componentwise as a consequence of Proposition 2 in Hallin et al. (2015), with

$$\tilde{Z}_{Tt} = T^{-1/2} \left(\frac{t+1}{T} - \frac{1}{2} \right) U_{t,j}(\vartheta^{(T)}, \Sigma) J_m(\tilde{F}_p(\|\varepsilon_t(\vartheta^{(T)})\|_\Sigma)),$$

$$Z_{Tt} = T^{-1/2} \left(\frac{t+1}{T} - \frac{1}{2} \right) U_{t,j}(\vartheta, \Sigma) J_m(\tilde{F}_p(\|\varepsilon_t\|_\Sigma)),$$

$\tilde{P}_T = P_{\vartheta^{(T)}; \Sigma, f}^{(T)}$, $P_T = P_{\vartheta; \Sigma, f}^{(T)}$, and LR_{Tt} as in (C.21).

Note that Conditions (a)-(e) of Proposition 1 in Hallin et al. (2015), which are required also in Proposition 2 in Hallin et al. (2015), are satisfied (see Proposition A.2). As J_m is bounded and $|U_t(\theta, \Sigma)| \leq 1$ for all $\theta \in \Theta$, Condition (g) of Proposition 2 in Hallin et al. (2015) clearly holds; their Condition (h) follows straightforwardly.

In order to conclude, thus, we only have to establish that Condition (f) (same reference) holds true. We have

$$\begin{aligned} \sum_{t=1}^T E_{\vartheta; \Sigma, f} \left[\left(\tilde{Z}_{Tt}(LR_{Tt})^{1/2} - Z_{Tt} \right)^2 \mid \mathcal{F}_{T, t-1} \right] &\leq 2 \sum_{t=1}^T E_{\vartheta; \Sigma, f} \left[\left(\tilde{Z}_{Tt} - Z_{Tt} \right)^2 \mid \mathcal{F}_{T, t-1} \right] \\ &\quad + 2 \max_{t=1, \dots, T} \left((LR_{Tt})^{1/2} - 1 \right)^2 \sum_{t=1}^T E_{\vartheta; \Sigma, f} \left[\tilde{Z}_{Tt}^2 \mid \mathcal{F}_{T, t-1} \right]. \end{aligned}$$

It follows from (12) in Hallin et al. (2015) that $\max_{t=1, \dots, T} |(LR_{Tt})^{1/2} - 1| = o_P(1)$, and we already noted

$$\sum_{t=1}^T E_{\vartheta; \Sigma, f} \left[\tilde{Z}_{Tt}^2 \mid \mathcal{F}_{T, t-1} \right] = O_P(1).$$

Hence Condition (f) holds if we show

$$\lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T E_{\vartheta; \Sigma, f} \left(D_{t;m}^{(T)} \right)^2 = 0 \quad (\text{C.35})$$

with (writing ϵ_t for $\epsilon_t(\vartheta^{(T)})$ and U_t^ϵ for $U_t(\vartheta^{(T)}; \Sigma)$)

$$D_{t;m}^{(T)} := U_{t,j}^\epsilon J_m(\tilde{F}_p(\|\varepsilon_t\|_\Sigma)) - U_{t,j}^\epsilon J_m(\tilde{F}_p(\|\epsilon_t\|_\Sigma)).$$

Let $\eta > 0$ such that $\tilde{F}_p(\eta) < m^{-1}$ and note that $J_m(\tilde{F}_p(\|\varepsilon_t\|_\Sigma)) 1_{\{\|\varepsilon_t\|_\Sigma \leq \eta\}} = 0$. This yields, for $|D_{t;m}^{(T)}|$, the bound

$$|D_{t;m}^{(T)}| \leq \left| J_m(\tilde{F}_p(\|\varepsilon_t\|_\Sigma)) - J_m(\tilde{F}_p(\|\epsilon_t\|_\Sigma)) \right| + \left| J_m(\tilde{F}_p(\|\varepsilon_t\|_\Sigma)) \right| |U_{t,j}^\epsilon - U_{t,j}^\epsilon| 1_{\{\|\varepsilon_t\|_\Sigma > \eta\}}.$$

Uniform continuity of $J_m \circ \tilde{F}_p$ and (C.30) imply

$$\lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T E_{\vartheta; \Sigma, f} \left(J_m(\tilde{F}_p(\|\varepsilon_t\|_\Sigma)) - J_m(\tilde{F}_p(\|\epsilon_t\|_\Sigma)) \right)^2 = 0.$$

Similarly, (C.31) and bounded convergence yield

$$\frac{1}{T} \sum_{t=1}^T \mathbb{E}_{\vartheta; \Sigma, f} |U_{t,j}^\varepsilon - U_{t,j}^\epsilon|^2 1\{\|\varepsilon_t\|_\Sigma > \eta\} \leq \mathbb{E}_{\vartheta; \Sigma, f} \max_{t=1, \dots, T} |U_{t,j}^\varepsilon - U_{t,j}^\epsilon|^2 1\{\|\varepsilon_t\|_\Sigma > \eta\} = o(1).$$

Convergence in (C.35) follows, hence also Condition (f). Proposition 2 in Hallin et al. (2015) thus applies, which establishes (C.34). Noting that

$$\mathbb{E}_{\vartheta; \Sigma, f} [Z_{Tt}(h'_T \Delta_{\vartheta; \Sigma, f}) \mid \mathcal{F}_{t-1}] = \frac{1}{p} \mathcal{I}_{p,m}(f, g) \left(\frac{t}{T+1} - \frac{1}{2} \right) v_{Tt,j}$$

completes the proof of Part (iv) of Proposition 4.2. \square

C.5.5. Proof of Part (v)

The proof of Part (v) is classical, though the various rates of convergence involved make it more complicated than usual. For given $M > 0$, let $\Theta^{(T)}(M)$ denote the set of possible values of $\vartheta^{(T)}$, as described in the last part of Assumption 5. Observe that the set of sequences $t^{(1)}, t^{(2)}, \dots$ such that $t^{(T)} \in \Theta^{(T)}(M)$ for all $T \in \mathbb{N}$ is a countable set, and that each of those sequences constitutes a sequence of perturbations in the sense of (4.5). The proof of Lemma 4.4 in Kreiss (1987) then applies without changes. The result follows. \square

C.6. Proof of Proposition 4.3

Part (i) follows immediately from Proposition 4.2(ii). Part (ii) follows from an application of Le Cam's third lemma as in Proposition 4.2(iii). Note that, without loss of generality, we may assume $B = 0$, as the score associated with B depends on $\alpha' W_\phi$, which is independent of $[(I_p - P_\alpha) \alpha_\perp]' W_\phi$. Thus, applying Le Cam's third lemma as in Theorem 6.6 in Van der Vaart (2000) and using Girsanov's theorem, leads to the same distribution under the null as under local alternatives generated by B since the sharp bracket of the score for B and $[(I_p - P_\alpha) \alpha_\perp]' W_\phi$ vanishes. As a result, we obtain, for the asymptotic mean of $[(I_p - P_\alpha) \alpha_\perp]' \Sigma^{-1/2} \mathcal{S}_g^{(T)}(\vartheta; \Sigma)$, under $\mathbb{P}_{\vartheta^{(T)}; \Sigma; f}$,

$$\begin{aligned} & \frac{1}{12p} \mathcal{I}_p(f, g) \left| \beta_\perp \Psi_{\Gamma, \Pi}^{-1} \alpha'_\perp \mu \right|^2 [(I_p - P_\alpha) \alpha_\perp]' \Sigma^{-1} (\alpha b + \alpha_\perp d) \\ &= \frac{1}{12p} \mathcal{I}_p(f, g) \left| \beta_\perp \Psi_{\Gamma, \Pi}^{-1} \alpha'_\perp \mu \right|^2 [(I_p - P_\alpha) \alpha_\perp]' \Sigma^{-1} \alpha_\perp d. \end{aligned}$$

Part (iii), due to the premultiplication of $\mathcal{S}_g^{(T)}(\vartheta; \Sigma)$ by $[(I_p - P_\alpha) \alpha_\perp]' \Sigma^{-1/2}$, is a consequence of Proposition 4.2(iii)-(v). More precisely, the shift in (4.8) vanishes with respect to B as $[(I_p - P_\alpha) \alpha_\perp]' \Sigma^{-1} \alpha = 0$.

Shifts due to d and D (as in Appendix A) are not possible, as a cointegration rank r_0 is imposed as a constraint in the construction of the estimator. Concerning the remaining local parameters m , G , A , and b , premultiplying their shifts by $[(I_p - P_\alpha) \alpha_\perp]' \Sigma^{-1/2}$ yields zero as well. Finally, Part (iv) follows from the standard construction of (conditionally) most stringent tests in LAN experiments with nuisance parameters (Corollary 3.2 and the discussion thereafter) and the observation that, in view of Part (iii), equally stringent tests can be constructed when Σ and B are unknown. \square

D. Additional Monte Carlo results

This section presents the results of the Monte-Carlo study, which, for lack of space, could not be included in Section 5—where we refer to for notation and the description of data-generating mechanisms.

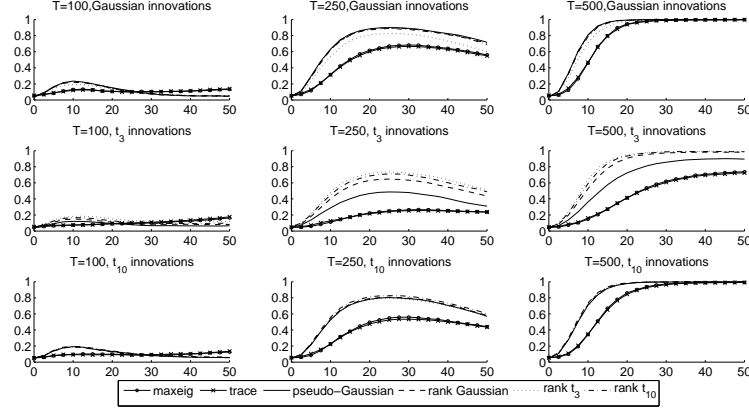
Table D.1: Simulated sizes (25,000 replications) of the *maxeig* test, *trace* test, $Q_{\dagger}^{(T)}$ and the rank-based tests (4.9), $g \in \{\phi, t_3, t_{10}\}$, under (5.1) for $p = 2$, $r_0 \in \{0, 1\}$, $\phi = -0.3$, $\Sigma \in \{I_2, \Sigma_{2,c}\}$, and $f \in \{\phi, t_3, t_{10}\}$. For $r_0 = 1$, *maxeig* and *trace* coincide.

	Sample size and innovation distribution								
	$T = 100$ and $\Sigma = I_2$			$T = 250$ and $\Sigma = I_2$			$T = 500$ and $\Sigma = I_2$		
Test	$f = \phi$	$f = t_3$	$f = t_{10}$	$f = \phi$	$f = t_3$	$f = t_{10}$	$f = \phi$	$f = t_3$	$f = t_{10}$
$r_0 = 0$									
<i>maxeig</i>	0.053	0.055	0.052	0.051	0.052	0.048	0.051	0.052	0.049
<i>trace</i>	0.055	0.057	0.054	0.053	0.056	0.049	0.052	0.052	0.050
$Q_{\dagger}^{(T)}$	0.047	0.044	0.044	0.051	0.047	0.048	0.049	0.048	0.048
$Q_{\phi}^{(T)}$	0.045	0.045	0.043	0.051	0.047	0.046	0.048	0.049	0.048
$Q_{t_3}^{(T)}$	0.050	0.048	0.048	0.051	0.048	0.047	0.048	0.048	0.047
$Q_{t_{10}}^{(T)}$	0.049	0.048	0.046	0.052	0.048	0.047	0.048	0.050	0.049
$r_0 = 1$									
<i>maxeig</i>	0.050	0.057	0.051	0.051	0.054	0.049	0.051	0.052	0.049
$Q_{\dagger}^{(T)}$	0.041	0.037	0.039	0.047	0.048	0.046	0.048	0.049	0.048
$Q_{\phi}^{(T)}$	0.039	0.038	0.036	0.047	0.047	0.045	0.047	0.048	0.047
$Q_{t_3}^{(T)}$	0.042	0.045	0.043	0.048	0.048	0.045	0.049	0.050	0.047
$Q_{t_{10}}^{(T)}$	0.041	0.042	0.040	0.047	0.048	0.044	0.049	0.050	0.048
	$T = 100$ and $\Sigma = \Sigma_{2,c}$			$T = 250$ and $\Sigma = \Sigma_{2,c}$			$T = 500$ and $\Sigma = \Sigma_{2,c}$		
Test	$f = \phi$	$f = t_3$	$f = t_{10}$	$f = \phi$	$f = t_3$	$f = t_{10}$	$f = \phi$	$f = t_3$	$f = t_{10}$
$r_0 = 0$									
<i>maxeig</i>	0.052	0.055	0.052	0.050	0.051	0.049	0.050	0.053	0.050
<i>trace</i>	0.055	0.056	0.055	0.052	0.052	0.049	0.051	0.053	0.049
$Q_{\dagger}^{(T)}$	0.047	0.040	0.044	0.051	0.047	0.048	0.049	0.048	0.048
$Q_{\phi}^{(T)}$	0.045	0.045	0.043	0.051	0.047	0.046	0.048	0.049	0.048
$Q_{t_3}^{(T)}$	0.050	0.048	0.048	0.051	0.048	0.047	0.048	0.048	0.047
$Q_{t_{10}}^{(T)}$	0.049	0.048	0.046	0.052	0.048	0.047	0.048	0.050	0.049
$r_0 = 1$									
<i>maxeig</i>	0.069	0.076	0.067	0.061	0.060	0.059	0.054	0.055	0.051
$Q_{\dagger}^{(T)}$	0.047	0.044	0.044	0.052	0.048	0.050	0.048	0.048	0.048
$Q_{\phi}^{(T)}$	0.042	0.045	0.041	0.050	0.049	0.048	0.048	0.048	0.047
$Q_{t_3}^{(T)}$	0.048	0.053	0.048	0.049	0.052	0.048	0.049	0.051	0.050
$Q_{t_{10}}^{(T)}$	0.047	0.049	0.045	0.051	0.052	0.048	0.049	0.049	0.048

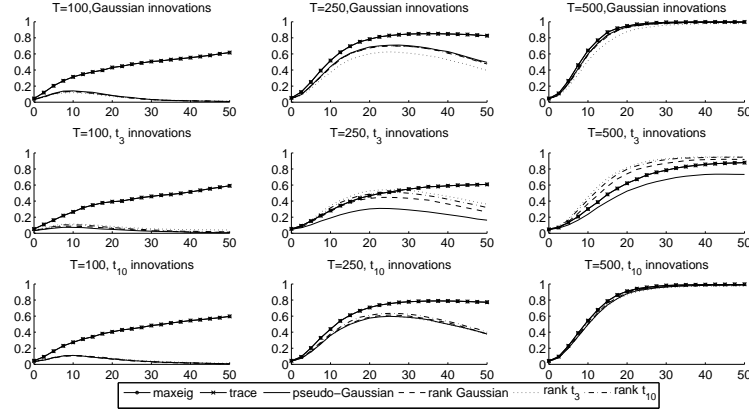
Table D.2: Simulated sizes (25,000 replications) of the *maxeig* test, *trace* test, $Q_{\dagger}^{(T)}$ and the rank-based tests (4.9), $g \in \{\phi, t_3, t_{10}\}$, under (5.1) for $p = 3$, $r_0 \in \{0, 1, 2\}$, $\phi = -0.3$, $\Sigma \in \{I_3, \Sigma_{3,c}\}$, and $f \in \{\phi, t_3, t_{10}\}$. For $r_0 = 2$, *maxeig* and *trace* coincide.

Sample size and innovation distribution									
Test	$T = 100$ and $\Sigma = I_3$			$T = 250$ and $\Sigma = I_3$			$T = 500$ and $\Sigma = I_3$		
	$f = \phi$	$f = t_3$	$f = t_{10}$	$f = \phi$	$f = t_3$	$f = t_{10}$	$f = \phi$	$f = t_3$	$f = t_{10}$
$r_0 = 0$									
<i>maxeig</i>	0.054	0.064	0.056	0.054	0.056	0.053	0.050	0.055	0.052
<i>trace</i>	0.057	0.064	0.058	0.054	0.056	0.053	0.050	0.054	0.051
$Q_{\dagger}^{(T)}$	0.046	0.041	0.044	0.048	0.043	0.051	0.048	0.048	0.050
$Q_{\phi}^{(T)}$	0.044	0.042	0.043	0.048	0.045	0.050	0.047	0.048	0.050
$Q_{t_3}^{(T)}$	0.047	0.045	0.046	0.049	0.046	0.051	0.050	0.047	0.050
$Q_{t_{10}}^{(T)}$	0.047	0.045	0.045	0.049	0.046	0.052	0.048	0.048	0.050
$r_0 = 1$									
<i>maxeig</i>	0.042	0.042	0.039	0.052	0.050	0.052	0.052	0.050	0.051
<i>trace</i>	0.042	0.045	0.041	0.051	0.052	0.051	0.050	0.052	0.052
$Q_{\dagger}^{(T)}$	0.034	0.030	0.034	0.045	0.041	0.047	0.048	0.046	0.048
$Q_{\phi}^{(T)}$	0.032	0.033	0.032	0.044	0.042	0.047	0.047	0.046	0.047
$Q_{t_3}^{(T)}$	0.036	0.041	0.040	0.047	0.046	0.048	0.048	0.050	0.048
$Q_{t_{10}}^{(T)}$	0.034	0.037	0.036	0.046	0.045	0.048	0.048	0.047	0.049
$r_0 = 2$									
<i>maxeig</i>	0.045	0.050	0.047	0.049	0.048	0.049	0.049	0.050	0.049
$Q_{\dagger}^{(T)}$	0.034	0.031	0.035	0.045	0.043	0.045	0.047	0.048	0.048
$Q_{\phi}^{(T)}$	0.031	0.032	0.033	0.044	0.044	0.043	0.046	0.049	0.046
$Q_{t_3}^{(T)}$	0.037	0.040	0.040	0.046	0.046	0.045	0.049	0.051	0.047
$Q_{t_{10}}^{(T)}$	0.033	0.036	0.035	0.045	0.046	0.044	0.047	0.050	0.047
Test	$T = 100$ and $\Sigma = \Sigma_{3,c}$			$T = 250$ and $\Sigma = \Sigma_{3,c}$			$T = 500$ and $\Sigma = \Sigma_{3,c}$		
	$f = \phi$	$f = t_3$	$f = t_{10}$	$f = \phi$	$f = t_3$	$f = t_{10}$	$f = \phi$	$f = t_3$	$f = t_{10}$
$r_0 = 0$									
<i>maxeig</i>	0.054	0.065	0.056	0.055	0.056	0.053	0.050	0.056	0.052
<i>trace</i>	0.056	0.065	0.058	0.053	0.057	0.052	0.051	0.054	0.050
$Q_{\dagger}^{(T)}$	0.046	0.041	0.044	0.048	0.043	0.051	0.048	0.048	0.050
$Q_{\phi}^{(T)}$	0.044	0.042	0.043	0.048	0.045	0.050	0.047	0.048	0.050
$Q_{t_3}^{(T)}$	0.047	0.045	0.046	0.049	0.046	0.051	0.050	0.047	0.050
$Q_{t_{10}}^{(T)}$	0.047	0.045	0.045	0.049	0.046	0.052	0.048	0.048	0.050
$r_0 = 1$									
<i>maxeig</i>	0.066	0.067	0.067	0.058	0.057	0.058	0.054	0.052	0.054
<i>trace</i>	0.073	0.072	0.073	0.061	0.060	0.062	0.055	0.055	0.054
$Q_{\dagger}^{(T)}$	0.040	0.038	0.038	0.047	0.041	0.046	0.048	0.046	0.047
$Q_{\phi}^{(T)}$	0.039	0.044	0.037	0.045	0.044	0.045	0.047	0.047	0.047
$Q_{t_3}^{(T)}$	0.045	0.050	0.045	0.048	0.048	0.048	0.048	0.048	0.049
$Q_{t_{10}}^{(T)}$	0.042	0.048	0.041	0.047	0.047	0.047	0.048	0.049	0.049
$r_0 = 2$									
<i>maxeig</i>	0.069	0.073	0.069	0.057	0.058	0.058	0.054	0.056	0.053
$Q_{\dagger}^{(T)}$	0.040	0.039	0.039	0.047	0.047	0.048	0.050	0.049	0.048
$Q_{\phi}^{(T)}$	0.037	0.042	0.037	0.046	0.047	0.047	0.048	0.050	0.047
$Q_{t_3}^{(T)}$	0.044	0.052	0.045	0.048	0.052	0.049	0.050	0.052	0.050
$Q_{t_{10}}^{(T)}$	0.041	0.047	0.040	0.047	0.050	0.049	0.049	0.052	0.049

Figure D.1: Simulated (2,500 replications) finite-sample powers of the *maxeig* test, the *trace* test, the pseudo-Gaussian test and the rank-based tests (4.9), $g \in \{\phi, t_3, t_{10}\}$,



(a) for testing $H : r = 0$ versus $H' : r = 1$ under (5.2), for $h \in \{0, 2.5, 5, \dots, 50\}$, $p = 2$, $T \in \{100, 250, 500\}$, $\Sigma = I_2$, and $f \in \{\phi, t_3, t_{10}\}$;

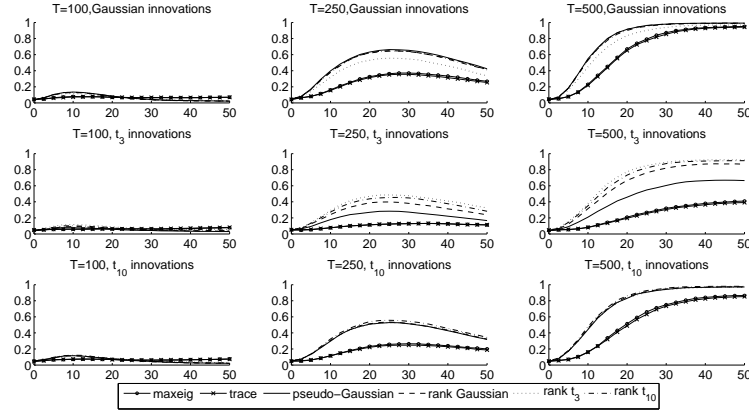


(b) for testing $H : r = 1$ versus $H' : r = 2$ under (5.2), for $h \in \{0, 2.5, 5, \dots, 50\}$, $p = 2$, $T \in \{100, 250, 500\}$, $\Sigma = I_2$, and $f \in \{\phi, t_3, t_{10}\}$.

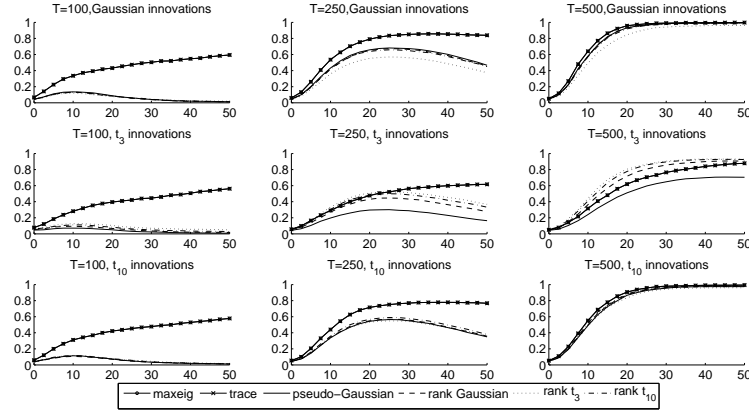
Table D.3: Simulated sizes (25,000 replications) of the *maxeig* test, *trace* test, $Q_{\dagger}^{(T)}$ and the rank-based tests (4.9), $g \in \{\phi, t_3, t_{10}\}$, under (5.1) for $p = 5$, $r_0 \in \{0, \dots, 4\}$, $\phi = -0.3$, $\Sigma = I_5$, and $f \in \{\phi, t_3, t_{10}\}$. For $r_0 = 4$, *maxeig* and *trace* coincide.

Test	Sample size and innovation distribution								
	$T = 100$			$T = 250$			$T = 500$		
	$f = \phi$	$f = t_3$	$f = t_{10}$	$f = \phi$	$f = t_3$	$f = t_{10}$	$f = \phi$	$f = t_3$	$f = t_{10}$
$r_0 = 0$									
<i>maxeig</i>	0.063	0.083	0.065	0.053	0.065	0.055	0.049	0.061	0.054
<i>trace</i>	0.070	0.082	0.072	0.058	0.063	0.058	0.054	0.059	0.055
$Q_{\dagger}^{(T)}$	0.044	0.038	0.043	0.047	0.047	0.047	0.048	0.047	0.048
$Q_{\phi}^{(T)}$	0.044	0.043	0.043	0.047	0.049	0.046	0.047	0.048	0.048
$Q_{t_3}^{(T)}$	0.045	0.042	0.046	0.050	0.047	0.047	0.048	0.048	0.049
$Q_{t_{10}}^{(T)}$	0.047	0.045	0.045	0.049	0.048	0.047	0.047	0.047	0.049
$r_0 = 1$									
<i>maxeig</i>	0.017	0.029	0.016	0.048	0.058	0.049	0.052	0.058	0.052
<i>trace</i>	0.029	0.037	0.029	0.050	0.057	0.051	0.050	0.057	0.054
$Q_{\dagger}^{(T)}$	0.025	0.023	0.023	0.043	0.042	0.041	0.046	0.045	0.047
$Q_{\phi}^{(T)}$	0.024	0.029	0.022	0.042	0.044	0.042	0.045	0.047	0.046
$Q_{t_3}^{(T)}$	0.030	0.037	0.030	0.043	0.047	0.044	0.047	0.048	0.047
$Q_{t_{10}}^{(T)}$	0.026	0.033	0.025	0.043	0.045	0.044	0.046	0.048	0.047
$r_0 = 2$									
<i>maxeig</i>	0.010	0.017	0.011	0.047	0.054	0.046	0.050	0.056	0.051
<i>trace</i>	0.017	0.026	0.019	0.049	0.055	0.048	0.051	0.055	0.051
$Q_{\dagger}^{(T)}$	0.018	0.018	0.017	0.039	0.039	0.040	0.043	0.043	0.046
$Q_{\phi}^{(T)}$	0.016	0.021	0.018	0.039	0.042	0.040	0.043	0.048	0.045
$Q_{t_3}^{(T)}$	0.025	0.033	0.027	0.042	0.047	0.043	0.044	0.050	0.047
$Q_{t_{10}}^{(T)}$	0.019	0.027	0.020	0.039	0.045	0.042	0.044	0.047	0.045
$r_0 = 3$									
<i>maxeig</i>	0.020	0.024	0.019	0.048	0.054	0.047	0.050	0.053	0.049
<i>trace</i>	0.021	0.027	0.021	0.049	0.054	0.048	0.049	0.054	0.049
$Q_{\dagger}^{(T)}$	0.017	0.018	0.018	0.039	0.038	0.038	0.042	0.045	0.046
$Q_{\phi}^{(T)}$	0.016	0.022	0.017	0.038	0.041	0.039	0.042	0.046	0.045
$Q_{t_3}^{(T)}$	0.025	0.036	0.027	0.041	0.048	0.042	0.046	0.049	0.045
$Q_{t_{10}}^{(T)}$	0.019	0.026	0.020	0.039	0.043	0.039	0.044	0.047	0.045
$r_0 = 4$									
<i>maxeig</i>	0.040	0.047	0.038	0.044	0.048	0.045	0.046	0.049	0.050
$Q_{\dagger}^{(T)}$	0.026	0.025	0.024	0.040	0.040	0.040	0.044	0.046	0.047
$Q_{\phi}^{(T)}$	0.024	0.026	0.023	0.039	0.042	0.039	0.043	0.047	0.047
$Q_{t_3}^{(T)}$	0.031	0.039	0.032	0.041	0.046	0.044	0.045	0.048	0.048
$Q_{t_{10}}^{(T)}$	0.026	0.031	0.026	0.038	0.044	0.042	0.045	0.048	0.047

Figure D.2: Simulated (2,500 replications) finite-sample powers of the *maxeig* test, the *trace* test, the pseudo-Gaussian test and the rank-based tests (4.9), $g \in \{\phi, t_3, t_{10}\}$,

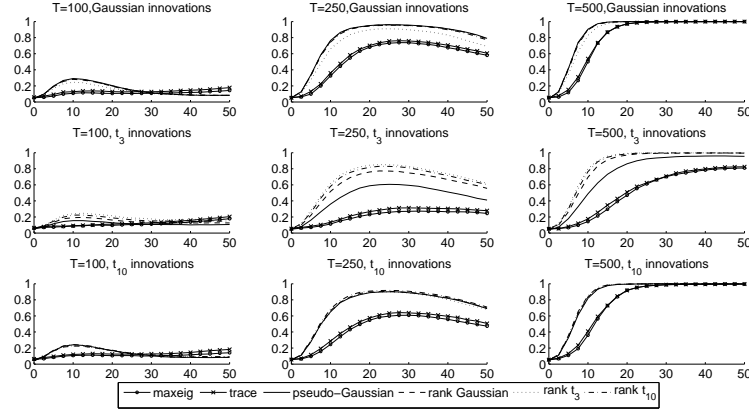


(a) for testing $H : r = 0$ versus $H' : r = 1$ under (5.2), for $h \in \{0, 2.5, 5, \dots, 50\}$, $p = 2$, $T \in \{100, 250, 500\}$, $\Sigma = \Sigma_{2,c}$, and $f \in \{\phi, t_3, t_{10}\}$;

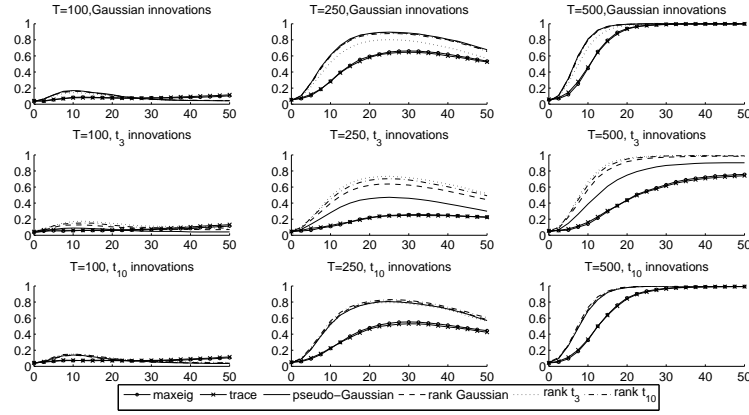


(b) for testing $H : r = 1$ versus $H' : r = 2$ under (5.2), for $h \in \{0, 2.5, 5, \dots, 50\}$, $p = 2$, $T \in \{100, 250, 500\}$, $\Sigma = \Sigma_{2,c}$, and $f \in \{\phi, t_3, t_{10}\}$.

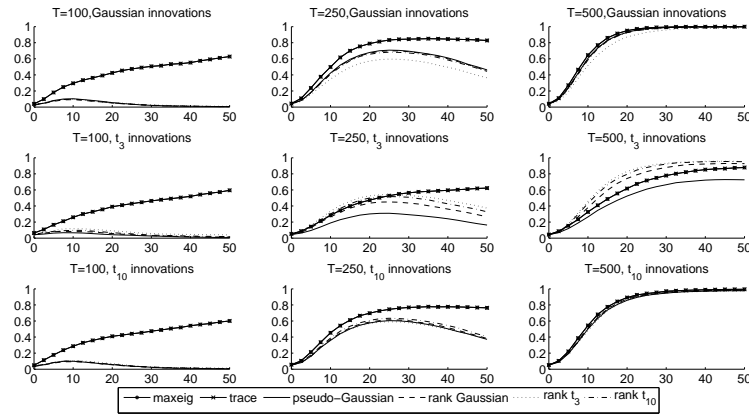
Figure D.3: Simulated (2,500 replications) finite-sample powers of the *maxeig* test, the *trace* test, the pseudo-Gaussian test and the rank-based tests (4.9), $g \in \{\phi, t_3, t_{10}\}$,



(a) for testing $H : r = 0$ versus $H' : r = 1$ under (5.2), for $h \in \{0, 2.5, 5, \dots, 50\}$, $p = 3$, $T \in \{100, 250, 500\}$, $\Sigma = I_3$, and $f \in \{\phi, t_3, t_{10}\}$;

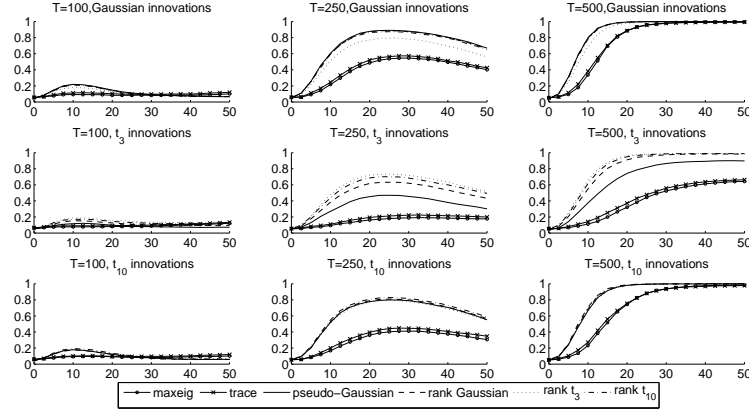


(b) for testing $H : r = 1$ versus $H' : r = 2$ under (5.2), for $h \in \{0, 2.5, 5, \dots, 50\}$, $p = 3$, $T \in \{100, 250, 500\}$, $\Sigma = I_3$, and $f \in \{\phi, t_3, t_{10}\}$;

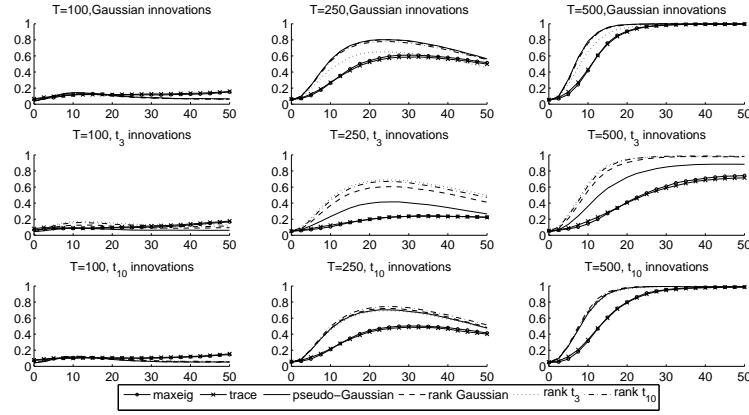


(c) for testing $H : r = 2$ versus $H' : r = 3$ under (5.2), for $h \in \{0, 2.5, 5, \dots, 50\}$, $p = 3$, $T \in \{100, 250, 500\}$, $\Sigma = I_3$, and $f \in \{\phi, t_3, t_{10}\}$.

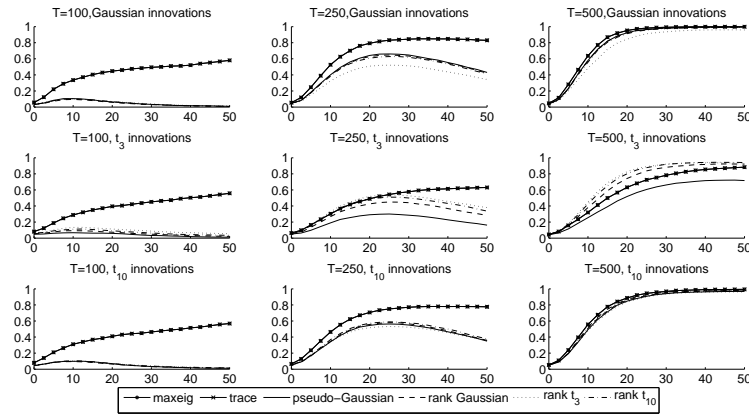
Figure D.4: Simulated (2,500 replications) finite-sample powers of the *maxeig* test, the *trace* test, the pseudo-Gaussian test and the rank-based tests (4.9), $g \in \{\phi, t_3, t_{10}\}$,



(a) for testing $H : r = 0$ versus $H' : r = 1$ under (5.2), for $h \in \{0, 2.5, 5, \dots, 50\}$, $p = 3$, $T \in \{100, 250, 500\}$, $\Sigma = \Sigma_{3,c}$, and $f \in \{\phi, t_3, t_{10}\}$;

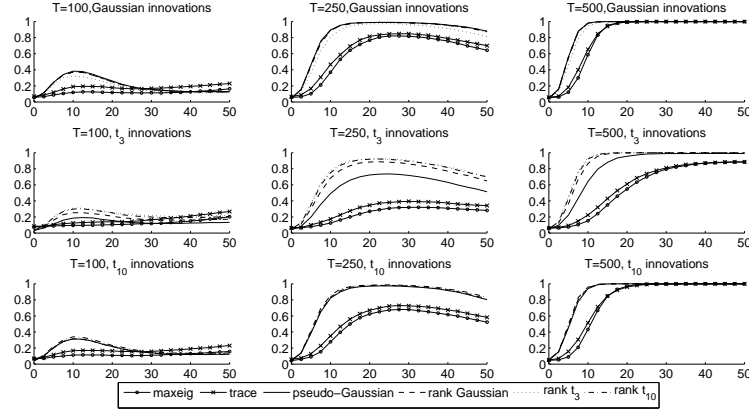


(b) for testing $H : r = 1$ versus $H' : r = 2$ under (5.2), for $h \in \{0, 2.5, 5, \dots, 50\}$, $p = 3$, $T \in \{100, 250, 500\}$, $\Sigma = \Sigma_{3,c}$, and $f \in \{\phi, t_3, t_{10}\}$;

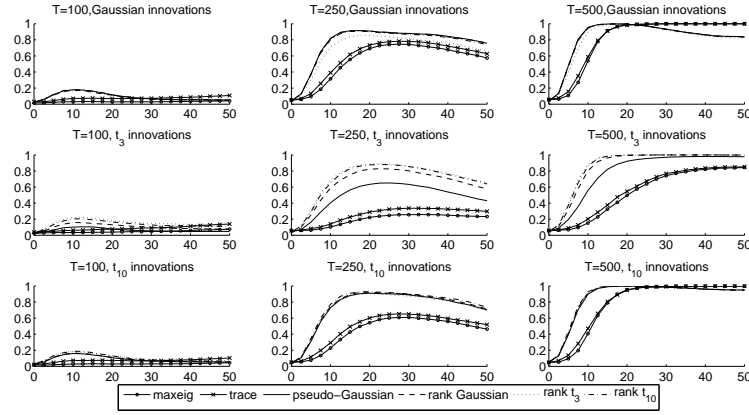


(c) for testing $H : r = 2$ versus $H' : r = 3$ under (5.2), for $h \in \{0, 2.5, 5, \dots, 50\}$, $p = 3$, $T \in \{100, 250, 500\}$, $\Sigma = \Sigma_{3,c}$, and $f \in \{\phi, t_3, t_{10}\}$.

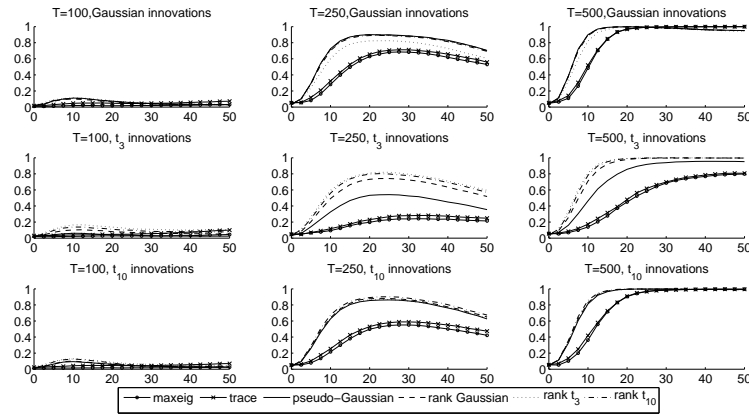
Figure D.5: Simulated (2,500 replications) finite-sample powers of the *maxeig* test, the *trace* test, the pseudo-Gaussian test and the rank-based tests (4.9), $g \in \{\phi, t_3, t_{10}\}$,



(a) for testing $H : r = 0$ versus $H' : r = 1$ under (5.2), for $h \in \{0, 2.5, 5, \dots, 50\}$, $p = 5$, $T \in \{100, 250, 500\}$, $\Sigma = I_5$, and $f \in \{\phi, t_3, t_{10}\}$;

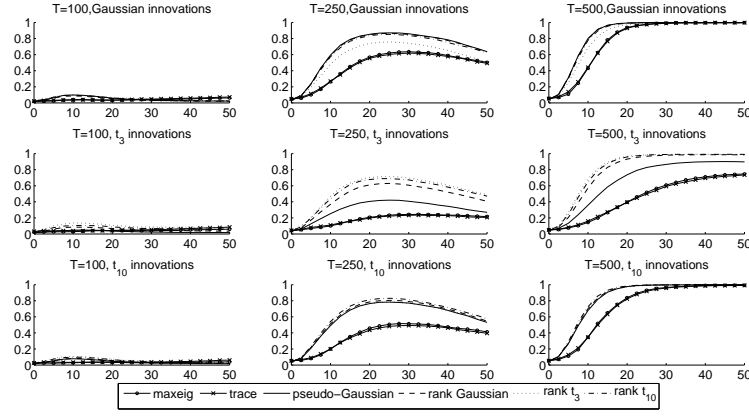


(b) for testing $H : r = 1$ versus $H' : r = 2$ under (5.2), for $h \in \{0, 2.5, 5, \dots, 50\}$, $p = 5$, $T \in \{100, 250, 500\}$, $\Sigma = I_5$, and $f \in \{\phi, t_3, t_{10}\}$;

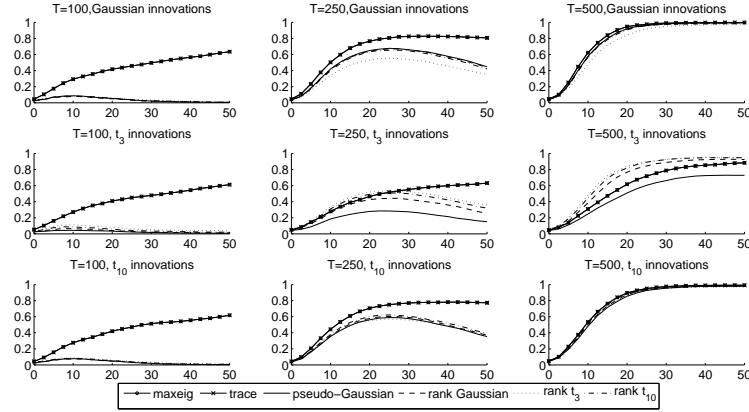


(c) for testing $H : r = 2$ versus $H' : r = 3$ under (5.2), for $h \in \{0, 2.5, 5, \dots, 50\}$, $p = 5$, $T \in \{100, 250, 500\}$, $\Sigma = I_5$, and $f \in \{\phi, t_3, t_{10}\}$.

Figure D.6: Simulated (2,500 replications) finite-sample powers of the *maxeig* test, the *trace* test, the pseudo-Gaussian test and the rank-based tests (4.9), $g \in \{\phi, t_3, t_{10}\}$,



(a) for testing $H : r = 3$ versus $H' : r = 4$ under (5.2), for $h \in \{0, 2.5, 5, \dots, 50\}$, $p = 5$, $T \in \{100, 250, 500\}$, $\Sigma = I_5$, and $f \in \{\phi, t_3, t_{10}\}$;



(b) for testing $H : r = 4$ versus $H' : r = 5$ under (5.2), for $h \in \{0, 2.5, 5, \dots, 50\}$, $p = 5$, $T \in \{100, 250, 500\}$, $\Sigma = I_5$, and $f \in \{\phi, t_3, t_{10}\}$.